Neural mechanisms supporting the effects of social context on economic decision making

Dissertation
Submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

To obtain the degree of
Doktor der Neuroökonomie, Dr. sc.
(corresponds to Doctor of Neuroeconomics, PhD)

presented by

Aidan Makwana
From the United Kingdom

Approved in February 2018 at the request of

Prof. Dr. Todd Hare
Prof. Dr. Philippe Tobler

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 14.02.2018

The Chairman of the Doctoral Board: Prof. Dr. Todd Hare

List of manuscripts

This dissertation is based on the following research articles:

Study 1:

Aidan Makwana, Georg Grön, Ernst Fehr and Todd Hare. A neural mechanism of strategic social choice under sanction-induced norm compliance (2015), eNeuro, May/June 2015, 2(3)/e0066-14.2015 1-8.

Study 2:

Aidan Makwana, Rafael Polania, Silvia Meier, Todd Hare. Model based brain substrates of fairness and money, *in preparation.*

Study 3:

Aidan Makwana, Ian Krajbich, Todd Hare. Multivariate classification of decision features between charitable and personal decisions, *in preparation.*

Acknowledgements

Over the course of my PhD program I have received a great deal of support, without which this dissertation would not have been possible. I would like to express my deepest gratitude to the following people.

First of all, to my supervisor Todd Hare. Without his patience and passion for this work, this thesis would not have been possible. Over the course of our many discussions I developed a deep appreciation for the nature of neuroeconomic research.

I would also like to thank the neuroecnomics faculty for their insight into my various projects including Philippe Tobler, Christian Ruff and Ernst Fehr. Their varied expertise and interests helped develop me into a more well rounded researcher.

The work presented here would also not have been possible without the work and help of my collaborators and I would like to thank Georg Grön, Ernst Fehr, Rafael Polania and Ian Krajbich for their input to the papers presented in this dissertation.

Throughout the course of this PhD program I have been fortunate to have colleagues that are both interested and skilled, as well as supportive of my research. To this end, I would like to thank Marcus Grüschow, Silvia Maier, Chris Burke, Tony Williams, Chaohui Guo, Azade Dogan, Friederike Meyer, Yosuke Morishima, Bastian Oud, Marius Moisa, Daniel Burghart, Ana Dereky, Gaia Lombardy, Andreas Mitsumasu, Anjali Raja-Behralle, Giuseppe Ugazio, Ana Cubillo, Chris Hill, Sebatian Weissengruber, Miguel Costa, Rachael Gwinn, Stephanie Smith James Wei and Arkady Konovalov. Their help and scientific insights also helped to form this thesis.

I addition to the scientific input, there are many others who helped in this work. In particular, I would like to thank the administrative and technical staff for their support, including Karl Treiber, Zoltan Nagy, Tamara Herz, Mirjam Britschgi, Mijam Bebi, Cornelia Schnyder and Sally Gschwend.

Finally and most importantly I would like to thank my family and Hope Sample for their constant support in completing this doctoral work.

Contents

Abstract

Decision making is a fundamental part of human life. Evidence from psychology, economics and more recently neuroscience, is providing new depths of understanding, from perceptual information acquisition to the effects of making decisions in social groups. This thesis uses a value based decision framework and incentive compatible experiments commonly found in experimental and behavioral economics to examine the effects of social phenomena on decision making. This makes use of current behavioral modeling methods to test the effects of other-perspective taking (mentalizing) and social norms on value based decisions.

In the first study, the value computations that take place in strategic social contexts were investigated where there was possibility of retribution for norm violations. Here, we used functional magnetic resonance imaging (fMRI) to show that when human subjects face such a context connectivity increases between the temporoparietal junction (TPJ), implicated in the representation of other peoples thoughts and intentions, and regions of ventromedial prefrontal cortex (vmPFC) that are associated with value computation. In contrast, we found no increase in connectivity between these regions in social nonstrategic cases where decision-makers are immune from retributive monetary punishments from a human partner. Moreover, there was also no increase in TPJ-vmPFC connectivity when the potential punishment was performed by a computer programmed to punish fairness norm violations in the same manner as a human would. Thus, TPJ-vmPFC connectivity is not simply a function of the social or norm enforcing nature of the decision, but rather occurs specifically in situations where subjects make decisions in a social context and strategically consider putative consequences imposed by others.

In the second study, the effects of social norms were investigated by applying the drift diffusion model (DDM) behavioral data from a modified ultimatum game that explored the effect of the social norm of fairness. It was found that the relative value of these decisions (as modeled in the DDM) could be influenced by focusing on the fairness or money in the decision. These decisions took place while the human subjects were undergoing fMRI scanning and the relative value correlated with large areas of medial frontal cortex. Given previous evidence showing the importance of these areas in

value computations, it appears that the DDM is able to effectively infer the value of the choices we presented. Further investigation revealed that the behavior was also reflected in some brain regions, with accepting unfair offers being associated with more activity in the anterior insula, while rejecting these unfair offers led to more activity in the TPJ. As such, it appears that different brain areas are involved with different behavior in the ultimatum game. Taken together, we show that the DDM is able to provide a good proxy for relative decision value across different behavior and in different contexts.

In the third study, the neural traces of value during purchase decisions was explored, examining common and distinct patterns between charitable and product decisions. These decisions were broken down into the willingness to pay and the price of a purchase and an integrated measure of choice value that depended on whether the option was chosen or not. This fMRI experiment also made use of a multivariate approach to examine patterns of activity in local areas of the brain. Testing where choice dependent integrated values were distinct between charity and product decisions revealed the dmPFC as discriminating value correlations between the two conditions. In addition to this choice dependant form of value, the different attributes of the purchase decision were also tested across local areas of the brain. These simple attributes were not seen to differ between charity and product purchase decisions. On the other hand, common encoding patterns appear to distinguish between the two attributes in the lateral prefrontal cortex and dorsal anterior cingulated. Thus, different areas of the brain encode different aspects of the decision, the choice dependent valuation being different across condition, but the simple value attributes as being common across the decision contexts.

In summary, the findings of these experiments emphasize the importance of social context in value based decision making. They hint at both common and distinct mechanisms that underlie these conditions and may help guide future research in making social contexts more amenable to economic analysis.

Chapter 1

General introduction

For humans, decision making is a fundamental part of living. As social animals, much of the decision making that humans are engaged in takes place in a social environment. This social context may be considered as being a variable that other decision variables are conditioned upon another example would be the contextual effect of the weather affecting the prospective enjoyment derived from outdoor vs. indoor activities. However, there appears to be something special about this social context. The high degree of sociability of humans and the complexity of the social environment combined with the apparent ease of integrating social information suggests that there may be structural qualities of the mind specially related to encoding social effects. This thesis explores some of these social effects on value based decision making by observing the human brain and behavior during decisions relating to social norms and value.

1.1 Social norms

One of the great questions in studying human behavior lies in understanding the nature of social norms. In general, normative behavior can assign value to an outcome without necessarily relying on self-interested utility calculations. For example, many people hold the moral norm that there is a value in the ethical treatment of animals, even when there is no direct benefit to themselves. Similarly, social norms can be seen as providing a similar form of value assignment in a social context. This is particularly evident in the

case of charitable donation, where giving anonymously to others may be done because it is considered to be simply the right thing to do. However, deciding the amount requires a form of value judgment. One way of formalizing charitable giving is by using an economic game known as the dictator game. In the standard variant of this game, two people take part, where one is given money and the opportunity to split it with the other player if they wish to (Kahneman et al. 1986). Even when steps are taken to anonymise the players and raise the stakes, players often share some of the money and even split the money equally in some cases. In this case, it appears that a social norm of fairness is influencing the choice behavior. The amount of influence this norm exerts can be calculated by examining how much money is given up in order to make the split more fair. Experimental evidence suggests that most people do give something, with fewer people giving larger fractions of the initial endowment, indicative of a trade off between the money that they keep and the value of a fair outcome (C. Camerer 2003). It should also be noted that the decreasing trend does not hold at the fairest split of 50%, suggesting that the value of social norms can be subject to heuristic judgment (Bos et al. 1997). Interestingly, the application of social norms to value decisions also appears to be subject to the context of the decision and the available actions (List 2007).

1.2 Strategic play

In addition to the impact social norms have on simple valuation, they can also shape the environment in which we make decisions. This is because in many cases, the results of our own actions depend on the actions of other people which may be governed by social norms. An example of this derives from a simple extension of the dictator game, known as the ultimatum game (Gth et al. 1982). This follows the same steps as the dictator game on the part of the person making the split. However, in the ultimatum game, the person who was a passive receiver can now decide to reject the split in which case the money is taken away from both players, or accept the split with the same result as if they were playing in the dictator game. In the absence of social norms, we might expect the receiver to accept anything as being better than nothing, but if they put a value on fairness, then they may decide to sacrifice their money in order to ensure a fair outcome where neither player receives anything. The more value they put on fairness, the more they would need before they accept a split. If the person deciding on the split believes

this to be the case, then they should take the other persons social norms into account when deciding how much to give in a strategic manner. Experimental evidence shows that recievers do reject unfair splits, even when they stand to lose money and people deciding the split apparently take this into account and make more fair offers. Although it has been suggested that this behavior follows from an irrational emotional response (Koenigs and Tranel 2007), the proportion of rejections decreases as more is offered (C. Camerer 2003). This suggests that in a value based framework, the enforcement and conformity of social norms may better explain the effects of social norms on human behavior.

1.3 Social contexts

One of the challenges in using a value based framework to understand social norms is that there are rarely objective measures of value. Most valuations are learnt from external stimuli by individuals, meaning that the value of an individual decision depends on the person making it and the situation that they find themselves in. An approach that has found much success in economics in to get people to reveal their preferences through the choices that they make (Samuelson 1938). However, in cases such as social contexts, the way that preferences are expressed may be different depending on the particular circumstance. For example, at the supermarket checkout, a person pushing into the queue may cause those who were already in line to punish the one who has pushed in. On the other hand, if there is someone behind in the queue with very few items, that person may even be invited to push in. Thus, the deployment of social norms may have radically different effects depending on the situation. Making inferences relating to these internal mechanisms from behavioral observation as in the case of a revealed preference can be difficult because the observations underdetermine the potential causes of the behavior. Ideally, a direct measure of an individuals value would provide a more firm basis for inferring the effects of social norms within a value based framework. In recent years, great progress has been made in relating the biological processes in the brain to behavior and the cognitive processes that may be driving it. Current research on neural correlates of value and environmental conditions that affect valuations in the field of neuroeconomics has proved fruitful. This research can guide our understanding of internal mechanisms that underpin behavior by showing whether a given hypothesis is supported by correlates in the brain.

1.4 Neural correlates of value

Some of the foundational work on a neural relationships between the brain and value based decision making came from the psychological research into reinforcement learning. This took work from Pavlov (Pavlov (1927) 2010) and tried to explain some variations on the original experiments such as operant conditioning (Skinner 1938) via a model of predicted outcomes compared to realized outcomes (Rescorla and Wagner 1972; Sutton and Barto 1998). In these models, a prediction error provides a signal for learning the structure of the world. During a learning experiment where macaque striatal neurons were recorded (Schultz 1998), it was found that firing patterns directly relating to this prediction error signal were present in dopaminergic neurons. Further experimentation has revealed that these results may even explain some discrepancies in the behavioral results from the theory, for example between losses and gains (Glimcher et al. 2008). Part of the power of this error signal is that it can relate to a general representation of expected value, rendering it amenable to economic questions. While these are normally framed in complex environments for humans, the advent of functional magnetic resonance imaging (fMRI) has allowed for the non-invasive measurement of the blood oxygen level dependant (BOLD) signal which may be considered as an indirect measure of neural activity. For example, an experiment examining simple purchasing choices found that the human equivalent of the striatum found in macaques, related to the prediction error signal (Knutson et al. 2007).

1.5 neural correlates of decision models

Prediction error is not the only model parameter that can be tracked and tested against neural correlates. For example, in a choice between two options, the drift diffusion model (DDM) takes the evidence in favor of each option and lets them compete against each other until a threshold is reached, whereupon a decision is made. This has proved very successful in capturing behavioral outcomes in perception tasks (Ratcliff and McKoon 2008) and more recently has been applied to value based decisions in different domains as well (Krajbich et al. 2015). The overall evidence in favor of an option and the evidence threshold have both been found to have neural correlates using a combined electroencephalographic (EEG) and fMRI experiment in a reinforcement learning setting (Frank et al. 2015). This modeling approach

captures the decision process itself as opposed to the feedback required in prediction error models. It can also be augmented to test the effects of experimentally manipulated variables (Polana et al. 2015) in order to capture the effects of different external conditions. In addition to the relative value of the decision options, tracking neural correlates may be used to examine the attributes that make up the value of an option. It has been shown that value representations that are made up from multiple attributes seem to be able to have those different attributes encoded in separate areas of the brain (Hare et al. 2009; Lebreton et al. 2009). However, there still appears to be a single area of the brain in ventral medial prefrontal cortex (vmPFC) that integrates these attributes into a single value (Lim et al. 2013). This structure of value formulation is important as it suggests that the integrated value of a choice may have a modular nature. In turn, this may yield parsimonious models of incorporating contextual information such as social norms into value based decision making.

1.6 The social brain

While the neural support for value based decision making may explain some social behaviors, there remains the question of how the social information is encoded in the first place. The precise nature of this social information is difficult to define, however one important function is the ability to understand the existence of other minds and be able to infer mental states from their actions. The study of this theory of mind (ToM) has shown that this ability develops in humans around the age of three to four years (Wellman et al. 2001). The universality of this human faculty suggests that there is a physiological basis, and indeed, recent work has shown a "mentalizing network of brain regions that are reliably activated during tasks where humans are asked to consider other peoples thoughts (Van Overwalle and Baetens 2009). Some areas of this network overlap with those often associated with value based decision making such as medial frontal cortex, while others appear to be more unique to the social information such as temporal parietal junction (TPJ) and superior temporal sulcus (STS). Recent work has shown that different elements of ToM seem to be related to these different regions, with information relating to value such as the confidence in social information being encoded in medial frontal cortex (Martino et al. 2017). It appears that this region may also be involved in emotional value in a social context, however, information about building beliefs about other minds being associated

with TPJ (Koster-Hale et al. 2017). In addition, charitable decisions appear to access this network in a similar way with social information being encoded in the TPJ and this region being coupled with the vmPFC when deciding donations (Hare et al. 2010). This evidence may provide an explanation for how information relating to social norms is integrated into value based decision making. On this view, the social information of others motivations and mental states encoded in the TPJ could be integrated with the understanding of social norms in order to provide an assessment of option value to the vmPFC. This thesis seeks to examine these effects over the course of three value based decision making fMRI experiments. These explore strategic play in a social norm compliance environment, the decision making process in fairness contexts and the similarities and differences in value representations during charitable and personal choices.

Chapter 2

Overview of studies

2.1 Study 1

Background

Social norm compliance is a key part of human sociality, where the applicability of a social norm must be considered to avoid peer punishment for norm violations. This requires taking your own actions and the likely reactions of other people into account. However, the way that value computations function in these contexts is poorly understood. As a central component of the brains decision circuitry, the vmPFC has been associated with value computation in non-strategic decision contexts ranging from primary to social rewards for both self and others (Nicolle et al. 2012; Bartra et al. 2013; Clithero and Rangel 2014) and in choices during competitive games (Hampton et al. 2008; Zhu et al. 2012). In addition, vmPFC lesions have been shown to alter choice behavior under strategic conditions where norm violations can result in retributive punishment (Krajbich et al. 2009). Collectively, these data suggest that vmPFC might compute subjective values in strategic social choices that require balancing personal preferences with predictions about how the reactions of others to norm violations will impact outcomes for self.

Previous research has shown that inferring another persons beliefs in order to estimate his probable future actions recruits neural circuits including the TPJ (Saxe and Wexler 2005; Frith and Frith 2006; Zhu et al. 2012). Moreover, studies on competitive and cooperative interpersonal games sug-

gest that TPJ encodes information about other players that could be used to guide choices (Behrens et al. 2008; Hampton et al. 2008; Coricelli and Nagel 2009; Bhatt et al. 2010; Hare et al. 2010; Rilling and Sanfey 2011; Carter et al. 2012; Morishima et al. 2012; Carter et al. 2012; Morishima et al. 2012). However, whether information encoded in TPJ is incorporated into vmPFC value signals during social norm enforcement choices is unknown.

Methods

Forty-seven healthy, right-handed male students performed a strategic economic game while undergoing fMRI scanning. The behavioral paradigm in the scanner consisted 24 trials. On all trials, participants split 100 monetary units (MUs) between themselves (Player A) and Player B. For 24 participants Player B represented a human counterpart and for 23 participants Player B was a computer. Subsequently, Player B made a decision about how many monetary units to spend on punishment (the computer was took actions based on previously acquired human behavioural data). The punishment rate selected by Player Bs decreased with greater transfers in an approximately linear fashion. All participants were randomly matched against different players on each trial (i.e. a one-shot game). Each trial consisted of a treatment screen indicating the trial type for 6s, a participant driven decision period (mean 4.3s, SD 2.7s), then a wait period of 6s followed by a feedback screen displayed for 6s. Trials were separated by a fixation cross ITI for 6 8.7s, sampled from a uniform distribution, thus the decision period started at least 12s after the previous trials feedback. During the task, participants faced 12 strategic trials where player A was punished 5 MUs for every MU that player B spent and 12 non-strategic trials, where player B could not punish player A at all. Both participants began every trial with a reserve of 25MUs and therefore, Player B was always able to punish Player A completely (i.e. take away all earnings) during the punishment trials. Our primary GLM for fMRI analysis modeled four regressor types: 1) treatment, 2) decision, 3) wait, and 4) feedback periods in all trials and separately for non-strategic only (8 regressor onsets in total). In addition, we used three parametric regressors (PR): PR1) kept amount at decision onset in all trials, PR2) kept amount at previous within-condition decision and PR3) profit amount at feedback onset for all trials. SPM 8 software was used to estimate this GLM and compute contrasts of interest in each individual participant. In addition, a second GLM was formulated to test the PPI effects relating to the vmPFC region correlating with value from the first analysis. The vmPFC time series was used as a physiological regressor and interacted with two separate psychological boxcar regressors for strategic and non-strategic trials.

This resulted in two separate PPI regressors. This PPI GLM consisted of the following nine regressors: 1) vmPFC time series, 2) non-strategic decision boxcar, 3) strategic decision boxcar, 4) non-strategic decision X vmPFC, 5) strategic decision X vmPFC, 6) non-strategic wait period (6 sec boxcar), 7) strategic wait period (6 sec boxcar), 8) non-strategic profit screen (6 sec boxcar), and 9) strategic profit screen (6 sec boxcar). Similar to the first GLM, parametric regressors for kept amount at decision, previous kept amount at decision and profit amount at feedback were included for both punishment and control conditions. Lastly, six motion parameter regressors were also included.

Results and Discussion

Behaviorally, there was no difference in total amounts transferred between the Social and Non-social treatment groups. The transfer rates in the non-strategic game are consistent with average rates (20 %) reported in the previous literature (C. Camerer 2003). Participants in the role of Player A transferred more in strategic than non-strategic conditions in both the social and the nonsocial treatments. These results suggest that Player A strategically increased the amount transferred to Player B to decrease the likelihood that Player B would exercise his punishment option and reduce Player As earnings regardless of whether Player B was a human or a computer programmed to mimic human reactions. In our initial neuroimaging analysis, we examined the degree to which vmPFC activity reflected value computations during monetary transfer decisions in both treatment types using a general linear model on blood oxygen level dependent (BOLD) signals. This analysis showed a positive association between kept amounts and vmPFC BOLD activity across all participants. The correlation between amount kept and BOLD activity in the vmPFC region of interest (ROI) was not significantly different between treatment groups indicating that participants playing against humans and computers represented the amount kept to an equal degree in vmPFC. The vmPFC result is consistent with theoretical models and existing empirical data suggesting a central role for vmPFC in the computation of subjective values for a wide range of decision contexts (Kable and Glimcher 2009; Hare et al. 2010; Rushworth et al. 2012; Bartra et al. 2013; Clithero and Rangel 2014). Such theories also posit that if vmPFC acts as a general valuation system, then its interactions will be modulated such that coupling with regions providing decision relevant information will increase. Next, we tested the hypothesis that the coupling between vmPFC and the right TPJ will increase more during decisions that require strategic evaluations of another persons response to the outcome than in complex-

ity matched control conditions using a psychophysiological interaction (PPI) analysis with the vmPFC as the seed region. We found that participants in the Social treatment showed more positive correlations between TPJ and vmPFC in strategic trials compared to the Non-social treatment. In order to test whether vmPFC-TPJ PPI strength is related to the overall strategic play of the participants, we tested whether the individual PPI difference contrast (strategic minus non-strategic) differentially correlated with participants average punishment amounts in the Social compared to Non-social groups. This second level, between subjects regression analysis revealed a link between vmPFC-TPJ PPI during strategic trials and average punishment levels that was stronger in Social than Non-social treatment participants. In the Social group, greater vmPFC-TPJ PPI was associated with less punishment by Player B, while there was no significant relationship in the Non-social group. Decisions that balance welfare for self with the impacts on and reactions of others to ones own choices are ubiquitous in social life. Our results provide insights into the neural mechanisms underlying such behavior and suggest a key role for interactions between TPJ and vmPFC. These findings are an important advance in our understanding of the neurobiology underlying strategic social choice and provide a basis for future investigations into this central aspect of human behavior.

2.2 Study 2

Background

The Ultimatum game (UG; Gth et al. 1982) is a widely used paradigm for studying social decisions in which monetary gain and fairness (i.e. equality) attributes both play important roles in determining players choices (see Methods; Fehr and Camerer 2007). The finding that people do not behave according to the Nash equilibrium (Mailath 1998) predicted by assumptions of purely selfish preferences in the UG is extremely robust (List 2007; Slonim and Roth 1998; Cameron 1999; Andersen et al. 2011). However, the precise proportion of the total amount proposers offer and responders are willing to accept varies across both individuals and choice contexts, indicating that expressed social preferences are malleable and at least partially state-dependent (Chang and Sanfey 2013; Wright et al. 2011; Sanfey 2009; Hoffman et al. 2000; Henrich et al. 2001; Andersen et al. 2011).

Previous UG experiments have shown that aspects of the UG choice sit-

uation relate to neural activity. For example, both objective and contextual aspects of fairness are reflected in the insula (Wright et al. 2011). Modifications to the UG can also give insight into the neural mechanisms underlying it, such as using the explicit cognitive strategy of reappraising the intentions of the proposer as more negative or positive that has been shown to recruit lateral prefrontal cortex (PFC) activity and to decrease and increase offer acceptance rates, respectively (Wout et al. 2010; Grecucci et al. 2013). Exogenously altering lateral PFC activity via transcranial magnetic stimulation (TMS) also changes the acceptance rate of unfair offers, despite leaving the fairness judgments of those offers unchanged (Knoch et al. 2006; Baumgartner et al. 2011).

Recent work has utilized sequential sampling models (SSMs) to examine social decision making processes and provided novel insights into interpersonal behavior. Originally applied to perceptual and categorization tasks, sequential sampling models have been extended to the domain of value-based choices (Forstmann et al. 2016), and more recently social decision making. Thus far, SSMs have been shown to account for behavior in both two-person asset allocation decisions (e.g. UG) and charitable donation decisions (Krajbich et al. 2015; Hutcherson et al. 2015). In fact, it has been shown that the parameters of an SSM fit to decisions over primary rewards for self in one sample of participants can accurately predict the choices and reaction times of a separate group of participants playing in the role of responder in the UG (Krajbich et al. 2015).

Methods

In total, 24 healthy, adults were entered into the analysis who had played our ultimatum game while undergoing fMRI scanning. This game consisted of two players: player A who was endowed with a sum of money and was able to offer a portion of this endowment to player B. Player B was then given the option to accept the offer or reject it, causing the whole endowment to be returned to the experimenter. The modification to the original Ultimatum game consisted of an attentional focus manipulation. Every offer was repeated over three focus conditions (Money, Fairness and Natural). The total endowment varied on each trial in order to minimize the correlation ($r = 0.40$) between monetary amount offered and the fairness of the ultimatums.

We analyzed participants decision processes in each condition using a hierarchical Drift Diffusion Model that allowed us to model both reaction times and choice outcomes. To this end, we estimated a drift diffusion model

for each condition, where the outcomes were accept or reject, and the starting bias and input function to the drift rate were allowed to vary. The input function was a linear combination of an intercept term, z-scored magnitude and fairness where the weights on each term were allowed to vary across conditions and subjects. Across subjects, the bias term and each of the weights on the input function to the drift rate was significantly greater than the null distribution. Similar to the logistic regression, there was a greater effect of offer magnitude in the money focus condition, and a greater effect of offer fairness in the fairness focus compared to the other conditions.

We used a standard mass univariate GLM approach to model these effects with a design matrix (GLM-1) where all trials were modeled as having the duration of reaction times. Thus the design matrix consisted of eight regressors: four regressors of interest: accept and reject onsets, with parametric modulators of offer and fairness aligned with each of these onsets and four regressors considered relevant to explain the variance in the experiment: trial onsets for fair and none conditions, as well as onsets for missed trials (duration 3 seconds) and the instruction screens (duration 5 seconds). GLM-1 was applied at a single subject level using SPM8. Group level non-parametric analysis was carried out using the FSL (Winkler et al. 2014) randomise function.

Using the DDM model fit parameters to derive an overall drift rate per trial, correlations with BOLD activity were tested using three GLMs where the durations were similar to those in GLM-1. In GLM-3, a dummy regressor and a parametric modulator of the drift rate was used for all trials. In addition, dummy regressors were included for the money fairness focus conditions as well as the rejected trials, missed trials and instruction screens in order to account for extra known sources of variance.

Results and discussion

Participants choice behavior was similar to that reported in previous studies using the ultimatum game (C. F. Camerer 2003; Fehr and Camerer 2007). The focus manipulations did change this, with more acceptance in the monetary condition and less acceptance in the fair condition compared to baseline. This suggests that our manipulation did elicit a behavioral effect in the direction that we had predicted.

The drift rate alone predicted subject responses. Comparisons of the coefficients reflecting the influence of offer magnitude and equality on choices in each condition revealed that offer magnitude had more impact during the

Money focus condition relative to both neutral focus and Baseline trials, while the influence of equality was greater when the initial focus was directed towards the fairness attribute.

The drift rate across all trials identified regions that reflected an integration of each trials monetary and fairness attributes because the rate of evidence accumulation (drift) in our DDM varies as a function of the trial-specific offer equality levels and magnitudes as well as the choice context (Baseline, Fairness or Money conditions). Multiple brain regions exhibited a BOLD signal that correlated positively with the regressor for drift rate, including the amygdala, dlPFC, dorsal and ventral portions of the medial PFC, fronto-polar cortex, insula (anterior and mid), precuneus, parietal cortex, thalamus, and ventral striatum. Next, we tested for brain regions that differentially represented the evidence for accepting the ultimatum as a function of the choice the participant ultimately made. This represents the drift towards the chosen choice and showed significant differences in regions such as vmPFC, striatum, and posterior cingulate cortex that have been shown to correlate with stimulus values for a wide range of goods and experiences (Bartra et al. 2013; Clithero and Rangel 2014) and to adapt value representations for multi-attribute stimuli according to context or attentional cues (Nicolle et al. 2012; Hare et al. 2011a; Rudorf and Hare 2014).

These results demonstrate that attentional cues can influence social decision making. and suggest that this effect can be captured by a DDM. This effect was expected given previous results that have changed the framing of objectively similar offers (Wright et al. 2011). Several studies have demonstrated that the perception of fairness can be decoupled from its effects on behavior This appears to be true whether the manipulation is based on framing (Wright et al. 2011), altering neurotransmitter levels (Crockett et al. 2008) or directly interfering with neural activity (Knoch et al. 2006). This suggests that there may be a difference between a more abstract knowledge of fairness and how it is evaluated and integrated into social decision making. In this study we observed the latter effect.

2.3 Study 3

Background

Much work has been done on the taxonomy of altruistic behavior in the

psychological literature (see Feigin et al. 2014, for a systematic review), differentiating principally between pure altruism (no net benefit to the self) and impure altruism (a net gain to the self as well as the other). These two types of altruism are then subdivided into several explanatory factors such as the role of emotion or expectations of reciprocity. The economic literature has taken a broadly similar path (Kagel and Roth 2016), but with more of an emphasis creating formal models to describe and explain the behavior (Hubbard et al. 2016; Gino et al. 2016). One of the most tractable forms of altruisms in a value based framework is charitable donation. This is often considered to be a pure altruism because the reputation and reciprocity effects can be controlled for, although it is possible that a 'warm glow' from charitable giving may deliver a primary reward.

Recent evidence suggests that interactions with wealth, goods and effort have different effects on the altruistic motive (Holmes et al. 2002; Strahilevitz 1999; Mayo and Tinsley 2009). One approach to shed light on the mechanisms underlying these interactions is to apply methods from neuroeconomics. Mounting evidence suggests that areas of the brain such as frontal cortex and striatum are vital to making value based decisions (Hare et al. 2010; Clithero and Rangel 2014). Several studies have used these methods in order to test whether there may be biological substrates of aspects of economic games such as the dictator game where there is an option to give money to another player who otherwise would not receive anything or charitable donations. For example (Izuma et al. 2009) found that being observed while deciding whether to donate to charity or keep the money for oneself increased charitable giving for difficult choices and that donating while under observation lead to a greater activation of the ventral striatum.

methods

Sixteen subjects were included in the fMRI study. Prior to scanning, subjects were told that they had received an endowment of 100 US dollars before starting the experiment. They then read the experiment instructions and completed a per-scan rating task in which they rated charities and household items by how much they were willing to pay (WTP) for them from their endowment. WTPs were elicited using a Becker de Groot auction. Subjects then read the instructions for the fMRI experiment and began 4 runs of the fMRI task while undergoing fMRI scanning.

Integrated decision value was tested in the brain using a GLM including a parametric regressor derived from taking the (WTP - price) conditional

on accepting the price (i.e. net value gained) and (price-WTP) conditional on refusing the item (i.e. net value lost). This regressor had onsets and durations aligned with the corresponding decision period as well as a boxcar regressor also with these onsets and durations. Six motion regressors were also included.

In addition, the different ways that decision features were encoded for each charity and purchase choice, were tested. This was achieved with a GLM including the main effect, WTP and price as well as the six motion regressors. As such, the design matrix was constructed with two dummy boxcar regressors for the charity and product conditions, with an onset for each purchase screen with duration equal to the reaction time. These GLMs were estimated on individuals' warped brains using SPM12.

These GLMs were used in mass univariate analyses and a cvMANOVA analysis, using a MANOVA framework cross validated using the 4 runs as separate folds for analysis. (Allefeld and Haynes 2014).

Results and discussion

While most behavior between conditions was not significantly different, the WTPs of charities and products were trend-level significant. However, a similar trend was observed in the pilot data and prices were chosen as to offset the impact of this difference. As such, during scanning, the average difficulty (defined as $|WTP - price|$) of product and charity trials was not significantly different.

The integrated decision value was tested against the BOLD signal using mass univariate and multivariate analyses. The univariate approach showed the BOLD signal in anterior cingulate cortex correlated negatively with the integrated decision variable, but only in product trials. Two analyses were tested using the multivariate approach First, the integrated decision value was tested across all trials, to examine whether cvMANOVA was sensitive to its effect on BOLD. Significant pattern discriminability was seen in large areas of the cortex, suggesting that locally distributed processing in these areas can track the effect of the integrated decision value. In addition, the difference in this decision value between charity and product trials was tested with cvMANOVA, revealing significant voxels in areas on the left lateral frontal cortex, demonstrating that this value may be encoded differentially in these brain areas depending on the type of decision made.

Testing the price and WTP separately, allowed these attributes and their

interactions to be tested against the BOLD data. Again, both univariate and multivariate analyses were tested. The main effects and differences did not show significant effects except for a negative correlation with price in the product condition alone. For the multivariate analysis, the voxels in the searchlight were predicted by two levels of condition and two levels of decision attribute. This decision attribute factor was tested against the BOLD signal across the brain and showed left lateral frontal cortex, although in different areas to the integrated decision result. One of the strengths of the MANOVA framework is that it allows cross decoding where there may be non-linear separation in the patterns exhibited. Thus, we were able to test the stability of the effect of the attribute factor while accounting for the effects of the condition factor, i.e. decoding the attribute encoding fit in one condition to test the other condition.

Chapter 3

General conclusions

3.1 Study 1

The first study investigated the neural mechanisms during a strategic and non-strategic game, playing against a human and against a computer. Strategic thinking can be difficult to precisely analyze, as different people may have different ways of processing the information and this may change even over the course of an experiment. Rather than try to develop a complete model of the decision process, the analysis of this study took a broader perspective and asked a particular question about BOLD activity instead. Specifically, was there a correlation with value in vmPFC in this strategic game, and was there a role for the TPJ relating to the presence of a social context. The correlation with vmPFC was found across all trials as might be expected if the vmPFC is playing an integrated role. However, it should be noted that the proxy for the expected value (amount kept) does not attempt to take into account the value associated with fairness, nor the amount they may have expected to be punished in the strategic condition. This may be an issue if the contribution of these sources of value are not proportional to the amount kept, which may be the case since the relationship with punishment and fairness may be non-linear, especially on the few occasions that more than an equal split was transferred. Additionally, the small amounts of variation in the non-strategic game may have made the estimates of correlations in the brain more variable. Despite this, there was still a correlation with the amount kept in the vmPFC, suggesting that the representation of the expected value is strong enough to outweigh the additional sources of variance.

One of the important negative findings was that there were no significant differences in the TPJ across conditions. This was unexpected considering the literature on the specific role of TPJ in a social setting. This may be related to the type of task often used in the theory of mind literature where subjects read a story before being asked to reflect on the mental states or on non-social aspects of the story (Gallagher et al. 2000). Playing a game with another player may already be enough to engage the mechanisms of the TPJ, even though the other player is a computer. One way to reconcile this effect is to consider the TPJ not as being specific to social contexts, but specific to the mental processes required in a social context, which may also be involved in other complex environments or the application of social norms. The psychophysiological interaction (PPI) effect between TPJ and vmPFC was more in line with the normal interpretation of its role, with very little change in coupling during a decion in the non-strategic condition and an increase in coupling in the social strategic condition. Another interesting point is that the coupling actually decreased in the non-social strategic condition. Coupled with there being no significant change in the activity in the TPJ, this may hint at some sort of context dependent information gating. Thus the TPJ may be engaged during non-social strategic games, but this information is not able to be integrated into the decision making process. The mechanism of this gating function is unknown, but a potential region would be the dlPFC, regulating the areas that receive the efferent TPJ neurons.

3.2 Study 2

The second study examined the decisions of the second player in an ultimatum game where subjects were encouraged to focus on different aspects of the decision. Several studies have examined neural activity during the ultimatum game, although these are often focused on the role of emotion and spite as opposed to the economic implications of the paradigm. Perhaps the clearest effect was that the personal gain representations of value up the dorsal frontal medial wall were strongest when subjects were asked to focus on the money. This result was not as clear when they were asked to focus on the fairness despite changes in behavior. Two issues may explain this; firstly the fraction offered may not be directly proportional to the perceived fairness of the offer considering that all of the offers were unfair Secondly, it is likely that correlations with fairness suffered from only having three levels of fairness to fit, although a simplified approach comparing the

most and least fair trials also did not show significant correlations. However, previous literature has demonstrated that different offers may be enough to effect behavioural changes, but not trigger reflective judgements of fairness . Thus it is possible that the unfairness of all offers means that the violations of the fairness social norm were fairly constant throughout the experiment. The importance of the drift rate correlations may in fact be larger than the results showed, considering that a similar approach using the probability of accepting the offer using a logistic regression with the same input function did not correlate with any neural activity. This may be related to the probability of choice capturing the simple behavior, whereas the drift rate may be closer to the actual neural mechanisms occurring during the decision. Although previous studies have shown the importance of drift rate in a social context (Krajbich et al. 2015), it is still surprising to see the strength of the effect, considering that the source of evidence is an abstract representation of money and fairness as opposed to a primary stimuli such as visual or audio shown in the psychophysics literature (Mulder et al. 2013). One of the results also suggests that the drift rate may be able to capture some more complex features of the decision process. When rejecting an offer, there was a positive correlation with the drift rate (in favor of accepting) in a portion of the dorsal cingulate. This may represent an effect similar to an effect of oddball trials in the stroop task (Veen and Carter 2002). The weaker the inclination to reject the offer, the more active this area was, suggesting that there may be some sort of conflict in the tradeoff when rejecting offers that is not present when accepting them.

3.3 Study 3

The third study used charitable donations compared to buying products in order to test contributing factors to altruistic decisions. Despite their importance in altruism, there are surprisingly few studies examining the neural mechanisms of charitable donation. This may be due to the difficulty in controlling all of the factors at play in altruistc choices. By including a comparison with purchasing for personal use, this study was able to get closer to the unique effects of benefiting others on the decision process. One surprising result in this experiment was how little BOLD activity correlated with either WTP or price during these decisions. While the nature of the task was to purchase an item, the integrated decision value was still expected to correlate with regions of the brain such as vmPFC. Since the prices were selected to

be close to the population average WTP, it is possible that hearing the price during the experiment acted as an anchor, altering their recall of the WTP and making each decision more difficult than the model would suggest. Subjects were significantly able to act in accordance with their WTP, so using price and WTP is still likely to be a good model to use, but perhaps there were more sources of noise contributing to a given voxel when calculating an integrated representation of value. This would explain why the searchlight analysis had more success at finding patterns in the data. With more sources to draw upon, it would be less susceptible to noise. In addition, both products and charities spanned a wide range of types. This may also explain why a single voxel was unable to fit an integrated value across all trials, whereas a searchlight had access to a larger area of the brain that was receiving constituent attributes of the value.

One of the more stringent analyses performed looked for areas of the brain where the BOLD pattern discriminant for differential decision attributes existed even when taking out the effects of the pattern discriminant for charity and product conditions. The region of anterior cingulate that survived this analysis represents an area where the tradeoff between price and WTP exists for value based decisions in both charity and personal purchase decisions.

3.4 General conclusions

Today, the social effects on neural substrates of value based decision making are an active area of research (Ruff and Fehr 2014). Considering the large role that sociability plays in our lives, this is a broad area of research and may help to explain cases where humans deviate from the simplified rationality assumptions made in economics. However, this is one piece of a greater puzzle of human economic behavior As our understanding of these effects increases, we may be able to integrate this with other important areas of economics, for example, how does mentalizing play a role in auctions. Social normative behavior could be integrated into game theory payoff matrices and whether they are deployed or not may determine the persistence of unstable coordination in the prisoners dilemma (Axelrod 2006). A more sophisticated model of charitable giving, including the many contributing factors may yield more accurate explanations of payments to public goods. Indeed, the integrating normative behavior more generally may even undermine some predictions of the standard economic model. For example, the efficient market hypothesis supposes that market prices are correct because people would exchange

goods until there was no self interested reason to continue (Fama 1970) One of the issues with this hypothesis is that people may deceive each other with regards to investment opportunities, but the prediction is that people will learn who can be trusted and to what extent through feedback. However, if the other-regarding social norms dictate behavior, then there may be no feedback to the person who instigated the investment, meaning that there might be no reason not to deceive.

In addition to the main psychological approaches presented here, the use of process models of decision such as the DDM may provide a mathematically amenable way to augment standard economic models of behavior This would dovetail with current economic thinking on using revealed preference, but would provide a more nuanced account of behavior than the current general axioms of revealed preference (Chambers et al. 2017). Although the frames of the experiments presented here were economic in nature, the key insights were in the developing understanding of neural mechanisms supporting behavior Across the three studies, methods such as connectivity analysis, model based fMRI and multivariate searchlight contributed to understanding in different ways.

Considering the spatial limitations of fMRI and the specificity of the cognitive information required for value based decision making, the question of how strong the inferences we can make is a difficult one to answer. On the one hand, to see such large scale neuronal effects in the volume of a voxel suggests that the effect must be strong. However, the diversity of neurons present in a given voxel may also contribute to that signal, which would suggest that the information carried in that voxel is less specific. Unfortunately fMRI cannot give detailed observations of neuronal activity and so it is necessary to bear in mind that inferences based on fMRI findings may not be as directly related to the experimental manipulations as they seem. Instead, fMRI provides an excellent method of forming the hypothesis space that we are willing to consider for explaining behavioral data.

In summary, the work presented here represents a small step towards a much greater problem. By showing that aspects of a value based framework correlate with neural activity, it adds weight to the argument that this framework should be used to understand social norms. This allows social information to be integrated into the wider decision making question rather than be treated as something distinct from simpler human behavior. I believe that the next steps in understanding how social norms are deployed taps into themes from judgment and decision making such as heuristics and learning.

In addition, I believe that future work in rigorous experiments on the effects of framing as integrated into a value based framework will be fruitful. Developing these formalizations of this very human behavior may even one day allow us to better understand the nature of intelligence and shape the development of artificial intelligence.

Bibliography

Allefeld, Carsten and John-Dylan Haynes (Apr. 2014). "Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA". In: *NeuroImage* 89, pp. 345–357. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.11.043. URL: http://www.sciencedirect.com/science/article/pii/S1053811913011920.

Andersen, Steffen et al. (2011). "Stakes Matter in Ultimatum Games". In: *The American Economic Review* 101.7, pp. 3427–3439. ISSN: 0002-8282. URL: http://www.jstor.org/stable/41408744.

Axelrod, Robert M. (Dec. 2006). *The Evolution of Cooperation: Revised Edition*. en. Google-Books-ID: KFf2HXzVO58C. Basic Books. ISBN: 978-0-465-00564-2.

Bartra, Oscar et al. (Aug. 2013). "The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value". In: *NeuroImage* 76, pp. 412–427. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.02.063. URL: http://www.sciencedirect.com/science/article/pii/S1053811913002188.

Baumgartner, Thomas et al. (Nov. 2011). "Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice". en. In: *Nature Neuroscience* 14.11, pp. 1468–1474. ISSN: 1097-6256. DOI: 10.1038/nn.2933. URL: http://www.nature.com/neuro/journal/v14/n11/full/nn.2933.html (visited on 07/28/2017).

Behrens, Timothy E. J. et al. (Nov. 2008). "Associative learning of social value". en. In: *Nature* 456.7219, pp. 245–249. ISSN: 0028-0836. DOI: 10.1038/nature07538. URL: http://www.nature.com/nature/journal/v456/n7219/full/nature07538.html?foxtrotcallback=true (visited on 09/14/2017).

Bhatt, Meghana A. et al. (Nov. 2010). "Neural signatures of strategic types in a two-person bargaining game". en. In: *Proceedings of the National Academy of Sciences* 107.46, pp. 19720–19725. ISSN: 0027-8424, 1091-

6490. DOI: 10.1073/pnas.1009625107. URL: http://www.pnas.org/content/107/46/19720 (visited on 09/14/2017).

Bos, K. van den et al. (May 1997). "How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect". eng. In: *Journal of Personality and Social Psychology* 72.5, pp. 1034–1046. ISSN: 0022-3514.

Camerer, Colin (2003). *Behavioral Game Theory : Experiments in Strategic Interaction*. en. Google-Books-ID: qzApnwEACAAJ. New Age International. ISBN: 978-81-224-3126-1.

Camerer, Colin F. (2003). "Ultimatum and Dictator Games: Basic Results". In: *Behavioral Game Theory: Experiments in Strategic Interaction*, pp. 48–58.

Cameron, Lisa A. (Jan. 1999). "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia". en. In: *Economic Inquiry* 37.1, pp. 47–59. ISSN: 1465-7295. DOI: 10.1111/j.1465-7295.1999.tb01415.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1465-7295.1999.tb01415.x/abstract.

Carter, R. McKell et al. (July 2012). "A distinct role of the temporal-parietal junction in predicting socially guided decisions". eng. In: *Science (New York, N.Y.)* 337.6090, pp. 109–111. ISSN: 1095-9203. DOI: 10.1126/science.1219681.

Chambers, Christopher P. et al. (May 2017). "General revealed preference theory". en. In: *Theoretical Economics* 12.2, pp. 493–511. ISSN: 1555-7561. DOI: 10.3982/TE1924. URL: http://onlinelibrary.wiley.com/doi/10.3982/TE1924/abstract.

Chang, Luke J. and Alan G. Sanfey (Mar. 2013). "Great expectations: neural computations underlying the use of social norms in decision-making". In: *Social Cognitive and Affective Neuroscience* 8.3, pp. 277–284. ISSN: 1749-5016. DOI: 10.1093/scan/nsr094. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3594719/.

Clithero, John A. and Antonio Rangel (Sept. 2014). "Informatic parcellation of the network involved in the computation of subjective value". In: *Social Cognitive and Affective Neuroscience* 9.9, pp. 1289–1302. ISSN: 1749-5016. DOI: 10.1093/scan/nst106. URL: https://academic.oup.com/scan/article/9/9/1289/1675099/Informatic-parcellation-of-the-network-involved-in (visited on 08/05/2017).

Coricelli, Giorgio and Rosemarie Nagel (June 2009). "Neural correlates of depth of strategic reasoning in medial prefrontal cortex". en. In: *Proceedings of the National Academy of Sciences* 106.23, pp. 9163–9168. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0807721106. URL: http://www.pnas.org/content/106/23/9163 (visited on 09/14/2017).

Crockett, Molly J. et al. (June 2008). "Serotonin modulates behavioral reactions to unfairness". In: *Science (New York, N.Y.)* 320.5884, p. 1739. ISSN: 0036-8075. DOI: 10.1126/science.1155577. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2504725/.

Fama, Eugene F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance* 25.2, pp. 383–417. ISSN: 0022-1082. DOI: 10.2307/2325486. URL: http://www.jstor.org/stable/2325486.

Fehr, Ernst and Colin F. Camerer (Oct. 2007). "Social neuroeconomics: the neural circuitry of social preferences". In: *Trends in Cognitive Sciences* 11.10, pp. 419–427. ISSN: 1364-6613. DOI: 10.1016/j.tics.2007.09.002. URL: http://www.sciencedirect.com/science/article/pii/S136466130700215X.

Feigin, Svetlana et al. (Oct. 2014). "Theories of human altruism: a systematic review". In: *Annals of Neuroscience and Psychology*. URL: http://www.vipoa.org/neuropsychol/1/1/ (visited on 08/30/2017).

Forstmann, B.U. et al. (2016). "Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions". In: *Annual review of psychology* 67, pp. 641–666. ISSN: 0066-4308. DOI: 10.1146/annurev-psych-122414-033645. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5112760/.

Frank, Michael J. et al. (Jan. 2015). "fMRI and EEG Predictors of Dynamic Decision Parameters during Human Reinforcement Learning". In: *The Journal of Neuroscience* 35.2, pp. 485–494. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.2036-14.2015. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4293405/.

Frith, Chris D. and Uta Frith (May 2006). "The Neural Basis of Mentalizing". In: *Neuron* 50.4, pp. 531–534. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2006.05.001. URL: http://www.sciencedirect.com/science/article/pii/S0896627306003448.

Gallagher, H. L. et al. (2000). "Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks". eng. In: *Neuropsychologia* 38.1, pp. 11–21. ISSN: 0028-3932.

Gino, Francesca et al. (2016). "Motivated Bayesians: Feeling Moral While Acting Egoistically". In: *Journal of Economic Perspectives* 30.3, pp. 189–212. URL: http://econpapers.repec.org/article/aeajecper/v_3a30_3ay_3a2016_3ai_3a3_3ap_3a189-212.htm.

Glimcher, Paul W. et al. (Oct. 2008). *Neuroeconomics: Decision Making and the Brain*. en. Google-Books-ID: g0QPLzBXDEMC. Academic Press. ISBN: 978-0-08-092106-8.

Grecucci, Alessandro et al. (Feb. 2013). "Reappraising the Ultimatum: an fMRI Study of Emotion Regulation and Decision Making". In: *Cerebral Cortex* 23.2, pp. 399–410. ISSN: 1047-3211. DOI: 10.1093/cercor/bhs028. URL: https://academic.oup.com/cercor/article/23/2/399/285843/Reappraising-the-Ultimatum-an-fMRI-Study-of (visited on 06/23/2017).

Gth, Werner et al. (Dec. 1982). "An experimental analysis of ultimatum bargaining". In: *Journal of Economic Behavior & Organization* 3.4, pp. 367–388. ISSN: 0167-2681. DOI: 10.1016/0167-2681(82)90011-7. URL: http://www.sciencedirect.com/science/article/pii/0167268182900117.

Hampton, Alan N. et al. (May 2008). "Neural correlates of mentalizing-related computations during strategic interactions in humans". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.18, pp. 6741–6746. ISSN: 1091-6490. DOI: 10.1073/pnas.0711099105.

Hare, Todd A. et al. (May 2009). "Self-control in decision-making involves modulation of the vmPFC valuation system". eng. In: *Science (New York, N.Y.)* 324.5927, pp. 646–648. ISSN: 1095-9203. DOI: 10.1126/science.1168450.

Hare, Todd A. et al. (Jan. 2010). "Value Computations in Ventral Medial Prefrontal Cortex during Charitable Decision Making Incorporate Input from Regions Involved in Social Cognition". en. In: *Journal of Neuroscience* 30.2, pp. 583–590. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4089-09.2010. URL: http://www.jneurosci.org/content/30/2/583 (visited on 06/23/2017).

Hare, Todd A. et al. (July 2011a). "Focusing Attention on the Health Aspects of Foods Changes Value Signals in vmPFC and Improves Dietary Choice". en. In: *Journal of Neuroscience* 31.30, pp. 11077–11087. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.6383-10.2011. URL: http://www.jneurosci.org/content/31/30/11077 (visited on 08/05/2017).

Henrich, Joseph et al. (2001). "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies". In: *The American Economic Review* 91.2, pp. 73–78. ISSN: 0002-8282. URL: http://www.jstor.org/stable/2677736.

Hoffman, Elizabeth et al. (June 2000). "The Impact of Exchange Context on the Activation of Equity in Ultimatum Games". en. In: *Experimental Economics* 3.1, pp. 5–9. ISSN: 1386-4157, 1573-6938. DOI: 10.1023/A:1009925123187. URL: https://link.springer.com/article/10.1023/A:1009925123187 (visited on 07/28/2017).

Holmes, John G. et al. (Mar. 2002). "Committing Altruism under the Cloak of Self-Interest: The Exchange Fiction". In: *Journal of Experimental Social Psychology* 38.2, pp. 144–151. ISSN: 0022-1031. DOI: 10.1006/jesp.2001.1494. URL: http://www.sciencedirect.com/science/article/pii/S0022103101914945.

Hubbard, Jason et al. (Oct. 2016). "A general benevolence dimension that links neural, psychological, economic, and life-span data on altruistic tendencies". eng. In: *Journal of Experimental Psychology. General* 145.10, pp. 1351–1358. ISSN: 1939-2222. DOI: 10.1037/xge0000209.

Hutcherson, Cendri A. et al. (July 2015). "A Neurocomputational Model of Altruistic Choice and Its Implications". English. In: *Neuron* 87.2, pp. 451–462. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2015.06.031. URL: http://www.cell.com/neuron/abstract/S0896-6273(15)00594-2 (visited on 07/28/2017).

Izuma, Keise et al. (Mar. 2009). "Processing of the Incentive for Social Approval in the Ventral Striatum during Charitable Donation". In: *Journal of Cognitive Neuroscience* 22.4, pp. 621–631. ISSN: 0898-929X. DOI: 10.1162/jocn.2009.21228. URL: http://dx.doi.org/10.1162/jocn.2009.21228.

Kable, Joseph W. and Paul W. Glimcher (Sept. 2009). "The Neurobiology of Decision: Consensus and Controversy". In: *Neuron* 63.6, pp. 733–745. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2009.09.003. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2765926/.

Kagel, John H. and Alvin E. Roth (Sept. 2016). *The Handbook of Experimental Economics, Volume 2: The Handbook of Experimental Economics*. en. Google-Books-ID: y4LRDAAAQBAJ. Princeton University Press. ISBN: 978-1-4008-8317-2.

Kahneman, Daniel et al. (1986). "Fairness and the Assumptions of Economics". In: *The Journal of Business* 59.4, S285–300. URL: http://econpapers.repec.org/article/ucpjnlbus/v_3a59_3ay_3a1986_3ai_3a4_3ap_3as285-300.htm.

Knoch, Daria et al. (Nov. 2006). "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex". en. In: *Science* 314.5800, pp. 829–832. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1129156. URL: http://science.sciencemag.org/content/314/5800/829 (visited on 07/28/2017).

Knutson, Brian et al. (Jan. 2007). "Neural predictors of purchases". In: *Neuron* 53.1, pp. 147–156. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2006.11.010. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1876732/.

Koenigs, Michael and Daniel Tranel (Jan. 2007). "Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27.4, pp. 951–956. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.4606-06.2007. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2490711/.

Koster-Hale, Jorie et al. (Nov. 2017). "Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs". In: *NeuroImage* 161.Supplement C, pp. 9–18. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2017.08.026. URL: http://www.sciencedirect.com/science/article/pii/S1053811917306730.

Krajbich, Ian et al. (Feb. 2009). "Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29.7, pp. 2188–2192. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.5086-08.2009.

Krajbich, Ian et al. (Oct. 2015). "A Common Mechanism Underlying Food Choice and Social Decisions". In: *PLOS Computational Biology* 11.10, e1004371. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004371. URL: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004371 (visited on 07/28/2017).

Lebreton, Mal et al. (Nov. 2009). "An automatic valuation system in the human brain: evidence from functional neuroimaging". eng. In: *Neuron* 64.3, pp. 431–439. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2009.09.040.

Lim, Seung-Lark et al. (May 2013). "Stimulus Value Signals in Ventromedial PFC Reflect the Integration of Attribute Value Signals Computed in Fusiform Gyrus and Posterior Superior Temporal Gyrus". en. In: *Journal of Neuroscience* 33.20, pp. 8729–8741. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4809-12.2013. URL: http://www.jneurosci.org/content/33/20/8729 (visited on 09/14/2017).

List, JohnA. (June 2007). "On the Interpretation of Giving in Dictator Games". In: *Journal of Political Economy* 115.3, pp. 482–493. ISSN: 0022-3808. DOI: 10.1086/519249. URL: http://www.journals.uchicago.edu/doi/abs/10.1086/519249.

Mailath, George (1998). "Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory". In: *Journal of Economic Literature* 36.3, pp. 1347–1374. URL: http://econpapers.repec.org/article/aeajeclit/v_3a36_3ay_3a1998_3ai_3a3_3ap_3a1347-1374.htm.

Martino, Benedetto De et al. (May 2017). "Social Information is Integrated into Value and Confidence Judgments According to its Reliability". en. In:

Journal of Neuroscience, pp. 3880–16. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3880-16.2017. URL: http://www.jneurosci.org/content/early/2017/05/31/JNEUROSCI.3880-16.2017 (visited on 09/29/2017).

Mayo, John W. and Catherine H. Tinsley (June 2009). "Warm glow and charitable giving: Why the wealthy do not give more to charity?" In: *Journal of Economic Psychology* 30.3, pp. 490–499. ISSN: 0167-4870. DOI: 10.1016/j.joep.2008.06.001. URL: http://www.sciencedirect.com/science/article/pii/S016748700800069X.

Morishima, Yosuke et al. (July 2012). "Linking Brain Structure and Activation in Temporoparietal Junction to Explain the Neurobiology of Human Altruism". In: *Neuron* 75.1, pp. 73–79. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2012.05.021. URL: http://www.sciencedirect.com/science/article/pii/S0896627312004874.

Mulder, Martijn J. et al. (July 2013). "The speed and accuracy of perceptual decisions in a random-tone pitch task". In: *Attention, Perception & Psychophysics* 75.5, pp. 1048–1058. ISSN: 1943-3921. DOI: 10.3758/s13414-013-0447-8. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3691469/.

Nicolle, Antoinette et al. (Sept. 2012). "An Agent Independent Axis for Executed and Modeled Choice in Medial Prefrontal Cortex". English. In: *Neuron* 75.6, pp. 1114–1121. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2012.07.023. URL: http://www.cell.com/neuron/abstract/S0896-6273(12)00674-5 (visited on 08/05/2017).

Pavlov (1927), P Ivan (July 2010). "Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex". In: *Annals of Neurosciences* 17.3, pp. 136–141. ISSN: 0972-7531. DOI: 10.5214/ans.0972-7531.1017309. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4116985/.

Polana, Rafael et al. (Aug. 2015). "The precision of value-based choices depends causally on fronto-parietal phase coupling". In: *Nature Communications* 6. ISSN: 2041-1723. DOI: 10.1038/ncomms9090. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4560799/.

Ratcliff, Roger and Gail McKoon (Apr. 2008). "The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks". In: *Neural computation* 20.4, pp. 873–922. ISSN: 0899-7667. DOI: 10.1162/neco.2008.12-06-420. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2474742/.

Rescorla, RA and Allan Wagner (Jan. 1972). "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonrein-

forcement". In: *Classical Conditioning II: Current Research and Theory*. Vol. Vol. 2.

Rilling, James K. and Alan G. Sanfey (2011). "The neuroscience of social decision-making". eng. In: *Annual Review of Psychology* 62, pp. 23–48. ISSN: 1545-2085. DOI: 10.1146/annurev.psych.121208.131647.

Rudorf, Sarah and Todd A. Hare (Nov. 2014). "Interactions between Dorsolateral and Ventromedial Prefrontal Cortex Underlie Context-Dependent Stimulus Valuation in Goal-Directed Choice". en. In: *Journal of Neuroscience* 34.48, pp. 15988–15996. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3192-14.2014. URL: http://www.jneurosci.org/content/34/48/15988 (visited on 08/05/2017).

Ruff, Christian C. and Ernst Fehr (Aug. 2014). "The neurobiology of rewards and values in social decision making". en. In: *Nature Reviews Neuroscience* 15.8, pp. 549–562. ISSN: 1471-003x. DOI: 10.1038/nrn3776. URL: https://www.nature.com/nrn/journal/v15/n8/box/nrn3776_BX4.html (visited on 09/14/2017).

Rushworth, Matthew F. S. et al. (Dec. 2012). "Valuation and decision-making in frontal cortex: one or many serial or parallel systems?" eng. In: *Current Opinion in Neurobiology* 22.6, pp. 946–955. ISSN: 1873-6882. DOI: 10.1016/j.conb.2012.04.011.

Samuelson, P. A. (1938). "A Note on the Pure Theory of Consumer's Behaviour". In: *Economica* 5.17, pp. 61–71. ISSN: 0013-0427. DOI: 10.2307/2548836. URL: http://www.jstor.org/stable/2548836.

Sanfey, Alan G. (June 2009). "Expectations and social decision-making: biasing effects of prior knowledge on Ultimatum responses". en. In: *Mind & Society* 8.1, pp. 93–107. ISSN: 1593-7879, 1860-1839. DOI: 10.1007/s11299-009-0053-6. URL: https://link.springer.com/article/10.1007/s11299-009-0053-6 (visited on 07/28/2017).

Saxe, Rebecca and Anna Wexler (2005). "Making sense of another mind: the role of the right temporo-parietal junction". eng. In: *Neuropsychologia* 43.10, pp. 1391–1399. ISSN: 0028-3932. DOI: 10.1016/j.neuropsychologia.2005.02.013.

Schultz, W. (July 1998). "Predictive reward signal of dopamine neurons". eng. In: *Journal of Neurophysiology* 80.1, pp. 1–27. ISSN: 0022-3077.

Skinner, Burrhus Frederic (1938). *The Behavior of Organisms: An Experimental Analysis*. en. Google-Books-ID: 13glAAAAMAAJ. Appleton-Century-Crofts.

Slonim, Robert and Alvin E. Roth (1998). "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic". In: *Econometrica* 66.3, pp. 569–596. ISSN: 0012-9682. DOI: 10.2307/2998575. URL: http://www.jstor.org/stable/2998575.

Strahilevitz, Michal (Jan. 1999). "The Effects of Product Type and Donation Magnitude on Willingness to Pay More for a Charity-Linked Brand". In: *Journal of Consumer Psychology*. Ethical Trade-Offs in Consumer Decision Making 8.3, pp. 215–241. ISSN: 1057-7408. DOI: 10.1207/s15327663jcp0803_02. URL: http://www.sciencedirect.com/science/article/pii/S1057740899703517.

Sutton, Richard S. and Andrew G. Barto (1998). *Reinforcement Learning: An Introduction*. en. Google-Books-ID: CAFR6IBF4xYC. MIT Press. ISBN: 978-0-262-19398-6.

Van Overwalle, Frank and Kris Baetens (Nov. 2009). "Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis". eng. In: *NeuroImage* 48.3, pp. 564–584. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2009.06.009.

Veen, Vincent van and Cameron S. Carter (Dec. 2002). "The anterior cingulate as a conflict monitor: fMRI and ERP studies". In: *Physiology & Behavior* 77.4, pp. 477–482. ISSN: 0031-9384. DOI: 10.1016/S0031-9384(02)00930-7. URL: http://www.sciencedirect.com/science/article/pii/S0031938402009307.

Wellman, Henry M. et al. (May 2001). "Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief". en. In: *Child Development* 72.3, pp. 655–684. ISSN: 1467-8624. DOI: 10.1111/1467-8624.00304. URL: http://onlinelibrary.wiley.com/doi/10.1111/1467-8624.00304/abstract.

Winkler, Anderson M. et al. (May 2014). "Permutation inference for the general linear model". In: *NeuroImage* 92, pp. 381–397. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2014.01.060. URL: http://www.sciencedirect.com/science/article/pii/S1053811914000913.

Wout, Mascha van t et al. (Dec. 2010). "The influence of emotion regulation on social interactive decision-making". In: *Emotion (Washington, D.C.)* 10.6, pp. 815–821. ISSN: 1528-3542. DOI: 10.1037/a0020069. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3057682/.

Wright, Nicholas D et al. (Apr. 2011). "Neural segregation of objective and contextual aspects of fairness". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.14, pp. 5244–5252. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.3138-10.2011. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3109551/.

Zhu, Lusha et al. (Jan. 2012). "Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning". en. In: *Proceedings of the National Academy of Sciences* 109.5, pp. 1419–1424. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1116783109. URL: http://www.pnas.org/content/109/5/1419 (visited on 09/14/2017).

Appendix A

Manuscript for study 1: Neural correlates of strategic play in a social context

eNeuro

New Research

Cognition and Behavior

A Neural Mechanism of Strategic Social Choice under Sanction-Induced Norm Compliance[1,2,3]

Aidan Makwana,[1] Georg Grön,[2] Ernst Fehr,[1] and Todd A. Hare[1]

DOI:http://dx.doi.org/10.1523/ENEURO.0066-14.2015

[1]Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, 8006 Zürich, Switzerland, and [2]University of Ulm, University Hospital for Psychiatry, 89075 Ulm, Germany

Abstract

In recent years, much has been learned about the representation of subjective value in simple, nonstrategic choices. However, a large fraction of our daily decisions are embedded in social interactions in which value guided decisions require balancing benefits for self against consequences imposed by others in response to our choices. Yet, despite their ubiquity, much less is known about how value computation takes place in strategic social contexts that include the possibility of retribution for norm violations. Here, we used functional magnetic resonance imaging (fMRI) to show that when human subjects face such a context connectivity increases between the temporoparietal junction (TPJ), implicated in the representation of other peoples' thoughts and intentions, and regions of ventromedial prefrontal cortex (vmPFC) that are associated with value computation. In contrast, we find no increase in connectivity between these regions in social nonstrategic cases where decision-makers are immune from retributive monetary punishments from a human partner. Moreover, there was also no increase in TPJ-vmPFC connectivity when the potential punishment was performed by a computer programmed to punish fairness norm violations in the same manner as a human would. Thus, TPJ-vmPFC connectivity is not simply a function of the social or norm enforcing nature of the decision, but rather occurs specifically in situations where subjects make decisions in a social context and strategically consider putative consequences imposed by others.

Key words: decision-making; fMRI; functional connectivity; norm compliance; strategy

Significance Statement

A large fraction of our decisions are embedded in social contexts that require balancing benefits for self against the positive or negative reactions of others in response to our choices. Yet, how the brain computes the value for different courses of action in such choices is unknown. We examined the neurobiological mechanisms underlying strategic social choices in the context of potential retributive punishment. Our findings indicate that there are specific increases in the functional interactions between brain regions previously associated with mentalizing about others' beliefs and key nodes of the brain's value computation system during choices in which it is necessary to balance direct personal gains against the likelihood of subsequent norm enforcing punishment by other people.

Introduction

A large portion of our daily decisions are embedded in social interactions in which the values of different behaviors depend on the behavior of relevant others. Such interactions range from major decisions about whether to apply for a new job and risk upsetting current colleagues to mundane choices about how much to tip the bartender

Received November 24, 2014; accepted May 28, 2015; First published June 16, 2015.

[1]The authors report no conflict of interest.

[2]Author contributions: This work represents a novel set of analyses on an existing dataset. G.G. and E.F. designed and performed the original research; A.M. and T.A.H. designed and conducted the current analyses; A.M., G.G., E.F., and T.A.H. wrote the paper.

[3]This work was supported in part by the Swiss National Science Foundation Grant 140277 to T.A.H. and the NCCR in Affective Sciences to E.F.

at your preferred pub in order to maintain favored patron status. In these and many other situations, social norm compliance must be considered to avoid peer punishment for norm violations. In all these cases, we need to take the likely reactions of other people into account. However, despite their ubiquity, very little is known about how value computation takes place in contexts where one's own behavior may trigger subsequent responses that affect subjective values.

As a central component of the brain's decision circuitry, the ventromedial prefrontal cortex (vmPFC) has been associated with value computation in nonstrategic decision contexts ranging from primary to social rewards for both self and others (Nicolle et al., 2012; Bartra et al., 2013; Clithero and Rangel, 2013) and in choices during competitive games (Hampton et al., 2008; Zhu et al., 2012). In addition, vmPFC lesions have been shown to alter choice behavior under strategic conditions where norm violations can result in retributive punishment (Krajbich et al., 2009). Collectively, these data suggest that vmPFC might compute subjective values in strategic social choices that require balancing personal preferences with predictions about how the reactions of others to norm violations will impact outcomes for self, but this idea has not yet been directly tested. Furthermore, how predictions about the opponents' behavior enter into vmPFC value computations is unknown. One hypothesis is that such information is provided to vmPFC by regions that are involved in mentalizing about others.

Previous research has shown that inferring another person's beliefs in order to estimate his probable future actions recruits neural circuits including the temporoparietal junction (TPJ; Saxe and Wexler, 2005; Frith and Frith, 2006; Zhu et al., 2012). Moreover, studies on competitive and cooperative interpersonal games suggest that TPJ encodes information about other players that could be used to guide choices (Behrens et al., 2008; Hampton et al., 2008; Coricelli and Nagel, 2009; Bhatt et al., 2010; Hare et al., 2010; Rilling and Sanfey, 2011; Carter et al., 2012; Morishima et al., 2012; Carter and Huettel, 2013). However, whether information encoded in TPJ is incorporated into vmPFC value signals during social norm enforcement choices is unknown. Therefore, we sought to examine whether TPJ-vmPFC interactions underlie value computations in this type of strategic social choice.

We examined brain activity using functional magnetic resonance imaging (fMRI) during decisions about the division of monetary assets between participants paired with either another human (social treatment) or a computer partner programmed to enforce social norm violations (nonsocial treatment). On each trial, participants had to choose how to divide 100 monetary units between themselves and the partner. However, these monetary allocation decisions were made in two distinct contexts. In the punishment context, the partner could punish perceived violations of the social norm for fairness by paying to reduce the participant's earnings, whereas in the control condition the partner could not enforce norm compliance through retributive punishment. The combination of these treatments and conditions allowed us to examine brain activity that was specific to choices that were both social and required strategic reasoning to optimize direct monetary gain against the probability of profit-reducing punishments for fairness norm violations.

Materials and Methods

Participants

Forty-seven healthy, right-handed male students performed a strategic economic game while undergoing fMRI scanning. Participants were screened for fMRI contraindications including acute medical conditions and psychiatric or neurological illness. All participants provided written informed consent in accordance with the local ethics committee.

Behavioral Paradigm

The behavioral paradigm proceeded as follows. On each trial, participants split 100 monetary units (MUs) between themselves (Player A) and Player B. For 24 participants Player B represented a human counterpart (social treatment group, mean age $\pm$ SD, 23.5 $\pm$ 2.3 years) and for 23 participants Player B was a computer (nonsocial treatment group, mean age $\pm$ SD, 24.8 $\pm$ 1.9 years). Participants were randomly assigned to either the social or nonsocial treatment groups upon arrival for the experiment. One participant from the social treatment was excluded from all analyses for a lack of comprehension of the task and two participants from the nonsocial group were excluded from the fMRI analyses described below because they never transferred any MUs (leaving 23 social and 21 nonsocial participants). The social group was instructed that each human Player B's punishment decisions had been acquired in a previous experiment using the strategy method. This method involved Player B making a decision about how many monetary units to spend on punishment if Player A transferred a specific amount. The punishment rate selected by human Player Bs decreased with greater transfers in an approximately linear fashion. The data from all Player Bs was used to generate a punishment distribution function and program the computer algorithm for the nonsocial treatment. The nonsocial group participants were instructed that they were playing against a computer that had been programmed to simulate the responses of the previous human Player B group and were given the same details as the social treatment participants about the strategy method of choice elicitation for Player Bs. All participants were randomly matched against different players on each trial (i.e. a one-shot game). Payment included 20 Euros for participating and 1 Euro per 100 MU earned. Each trial consisted of a treatment screen indicating the trial type for 6 s, a participant driven decision period (mean 4.3 s, SD 2.7 s), then a wait

Correspondence should be addressed to Todd A. Hare, Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zürich. E-mail: todd.hare@econ.uzh.ch.

DOI:http://dx.doi.org/10.1523/ENEURO.0066-14.2015

period of 6 s followed by a feedback screen displayed for 6 s. Trials were separated by a fixation cross ITI for 6–8.7 s, sampled from a uniform distribution, thus the decision period started at least 12 s after the previous trials feedback. During the task, participants faced 12 control trials (CON) and 12 punishment trials (PUN) in a random order as indicated during the treatment screen. In CON trials, Player B was not able to punish Player A for making a selfish split (i.e. a dictator game scenario); however, in PUN trials Player B could punish Player A by 5 MUs for each 1 MU spent. Both participants began every trial with a reserve of 25 MUs and therefore, Player B was always able to punish Player A completely (i.e. take away all earnings) during the punishment trials.

Behavioral Analysis

The behavioral variable of interest was the amount kept/transferred by participants in the role of Player A as a function of group and condition. There was a nonnormal distribution of transferred amounts in CON trials (Kolmogorov–Smirnoff test, $p = 0.03$), therefore, we analyzed the transfer amount data using nonparametric Wilcoxon signed rank (paired) and Kruskal–Wallis rank sum tests. All p values reported are based on two-sided tests. To better describe the punishment distributions, we linearly regressed punishment on the transfer amount for the social and nonsocial groups.

MRI Acquisition

Blood oxygen level-dependent (BOLD) echo planar imaging (EPI) scans were performed on a 3 Tesla Siemens Magnetom Allegra using 32 slices and a voxel resolution of $2 \times 2 \times 2$ mm (+0.5 mm slice gap), with a TR of 2490 ms, and a TE of 38 ms. All fMRI data was acquired during a single scanning session (mean length of 750 s, SD 38.5 s). A full brain EPI (56 slices using the same parameters as functional EPI) and anatomical scan (sagittal MPRAGE T1 sequence with a voxel size of $1 \times 1 \times 1$ mm) were also acquired. The fMRI data preprocessing included slice-time correction, spatial realignment to the mean EPI image for each subject, normalization to MNI space, and smoothing with a 10 mm FWHM Gaussian kernel using the SPM 8 software (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK).

fMRI Analysis

Our primary GLM (GLM-1) was computed to examine BOLD activity relating to the amount kept/transferred during the decision period. GLM-1 modeled four regressor types: (1) treatment, (2) decision, (3) wait, and (4) feedback periods in all trials (PUN and CON) and separately for CON only (8 regressor onsets in total). Single 0 s duration stick functions were convolved with the canonical HRF for the treatment, decision and feedback periods, and a 6 s boxcar function was used for convolution during the wait period. In addition, we used three parametric regressors (PR): (PR1) kept amount at decision onset in all trials, (PR2) kept amount at previous within-condition decision, and (PR3) profit amount at feedback onset for all trials. Six motion parameter regressors were also included in GLM-1. Note that the initial endowment is fixed at 100 MUs for every trial, and therefore, a positive correlation with the amount kept by Player A (PR1) implies a negative correlation with amount transferred to Player B.

SPM 8 software was used to estimate GLM-1 and compute contrasts of interest in each individual participant.

At the second level, we used the "randomise" function from the FSL 5.0.6 software package (http://www.fmrib.ox.ac.uk/fsl/) to test for regions that reflected the amount kept across all participants. We computed a one sample t-test on the single participant contrasts for positive correlations with the kept amount regressor together with a nuisance variable (0 = social, 1= nonsocial) to explain variance due to social and nonsocial participant groups. We performed the t test using the nonparametric permutation algorithm in randomise in combination with the threshold-free cluster enhancement (TFCE) method implemented in FSL (Smith and Nichols, 2009). Test statistics and p values were derived from 5000 permutations. We corrected for multiple comparisons using familywise error correction at the whole-brain level to achieve corrected significance levels of $p < 0.05$.

PPI Analysis

For each participant, a seed time course in vmPFC was extracted from a 4 mm sphere centered on the voxel with the strongest correlation with kept amount in that participant from within the overlapping voxels for the group vmPFC cluster generated by GLM-1 and an anatomical mask of vmPFC including the rectal gyrus, medial orbitofrontal, and anterior cingulate cortex below $z = 5$ (5464 8 mm^3 voxels) based on the AAL atlas (Tzourio-Mazoyer et al., 2002). The vmPFC time series was deconvolved as outlined by Gitelman et al., (2003) before creating the psychophysiological interaction regressors. For the psychophysiological interaction (PPI) GLM (GLM-PPI), the vmPFC time series was used as a physiological regressor and interacted with two separate psychological boxcar regressors for the decision period in CON and PUN conditions. This resulted in two separate psychophysiological interaction terms in GLM-PPI. In GLM-PPI, the decision period duration was modeled as 5 s before the first button press. This expanded window was used because the precise timing of the amount to keep/transfer computation within the treatment and decision screen periods cannot be determined in this task. However, this timing resolution limitation would not bias the results in favor of any specific decision type and, if anything, works against the current findings by adding noise to the analysis. GLM-PPI consisted of the following nine regressors: (1) vmPFC time series, (2) CON decision period boxcar, (3) PUN decision period boxcar, (4) CON decision $\times$ vmPFC, (5) PUN decision $\times$ vmPFC, (6) CON wait period (6 s boxcar), (7) PUN wait period (6 s boxcar), (8) CON profit screen (6 s boxcar), and (9) PUN profit screen (6 s boxcar). Note that, a one-way ANOVA for the SDs of the PPI regressors for group and condition showed that they were not significantly different ($F_{(1,83)} = 1.14$, $p = 0.338$) suggesting that the PPI analysis was not biased against CON conditions where kept amounts showed less variance. Similar

Table 1 Regions correlating with the amount kept by Player A at the time of choice

Region	Hemisphere	Extent	*x*	*y*	*z*	Peak T
Lingual gyrus	R/L	1257	8	−74	8	5.21
Cingulate gyrus	R/L	39	2	−10	36	4.54
vmPFCa-ACC	R/L	36	0	48	−2	4.14
mPFC-paracingulate gyrus	R/L	30	0	54	4	3.79
mPFC-ACC	R/L	28	−4	44	14	3.91
Frontopolar cortex/IFG	R	23	42	44	0	4.75
dmPFCb-paracingulate/SFG	R/L	21	−2	50	26	5.45
Occipital cortex	R	20	42	−76	−6	4.69
ACC	R/L	18	0	26	28	4.22
Thalamus	L	16	−12	−34	8	3.90
vmPFC-ACC	R	14	8	48	0	3.82
Frontopolar cortex	L	11	−16	58	28	5.00
Cingulate gyrus	L	10	−4	−4	32	3.95

Peak coordinates (*x*,*y*,*z*) are listed in MNI space. T values are test statistics derived from 5000 permutations of the data. All regions are significant at $p < 0.05$ whole-brain familywise error corrected for multiple comparisons.
IFG, Inferior frontal gyrus; SFG, superior frontal gyrus; R, right; L, left.
[a]vmPFC cluster used as a mask to extract subject specific time courses for PPI analyses.
[b]dmPFC cluster used as a mask to extract subject specific time courses for PPI analyses.

to GLM-1, parametric regressors for kept amount at decision, previous kept amount at decision and profit amount at feedback were included for both punishment and control conditions. Last, GLM-PPI included the six motion parameter regressors. A PPI analysis using the dorsomedial prefrontal cortex (dmPFC) seed noted in Table 1 was also performed. The analysis was identical to GLM-PPI, except that the BOLD time courses were extracted from the dmPFC ROI rather than the vmPFC ROI described above.

Following estimation of GLM-PPI in SPM8, single participant contrasts were computed for regressors of interest. At the second level, we again used TFCE and the nonparametric permutation function, randomise, to test for between group differences in connectivity with vmPFC. Test statistics and *p* values were derived from 5000 permutations. Based on previous work (Morishima et al., 2012) showing that social preferences during interpersonal interactions are linked to structural and functional differences in the TPJ, we created a spherical ROI with 10 mm radius around the MNI coordinates ($x, y, z = 60, -44, 18$). The conjunction of this ROI and the group functional coverage mask was used for small volume correction (324 8 mm^3 voxels). This functional coverage map was utilized because the acquisition parameters for the functional MRI data did not provide whole-brain coverage, and in some cases, the tilt of the transverse slices relative to anterior commissure–posterior commissure resulted in lack of coverage for the superior temporal and inferior parietal cortex. Forty-two participants (21 social and 21 nonsocial) had adequate functional coverage and were included in the PPI analysis. We corrected for multiple comparisons using familywise error correction within this mask to achieve small volume correction (SVC) of $p < 0.05$.

The bar plots shown in Figure 3*c* were created by taking the average vmPFC- TPJ PPI coefficients from all voxels in the functional ROI for the difference between social and nonsocial punishment trials shown in Figure 3*a*. These bar plots are presented for visualization purposes only and were not used as a basis for any statistical analysis.

In addition to comparing the PPIs during the PUN decisions between groups, we also tested for an association between the vmPFC-TPJ PPI during punishment decisions and the average punishment received by each individual within the social and nonsocial groups. We applied the same TPJ small volume correction described above for this analysis.

Last, we performed a *post hoc* analysis of correlations with profit during the PUN feedback condition (GLM-PPI regressor 9) by extracting PUN profit betas from all significant voxels in the social PUN PPI cluster shown in Figure 3*b*.

Results

Behaviorally, there was no difference in the total amounts transferred between the social and nonsocial treatment groups (Kruskal–Wallis $X^2_{(1,N=88)} = 0.48$, $p = 0.49$). Transfers in the social CON condition were on average 9.3 MU (SD 17.0), leading to an average percentage split of 22.9% for Player B after accounting for the 25 MU reserve amount for both players. These transfer rates are consistent with average rates (~20%) reported in the previous literature (Camerer, 2003). Participants in the role of Player A transferred more in PUN than CON conditions in both the social [Wilcoxon signed rank (W) = 276, $p = 2.88e{-}5$] and the nonsocial treatments (W = 231, $p = 6.36e{-}5$; Fig. 1). These results suggest that Player A strategically increased the amount transferred to Player B to decrease the likelihood that Player B would exercise his punishment option and reduce Player A's earnings regardless of whether Player B was a human or a computer programmed to mimic human reactions. Increasing the amount transferred in PUN trials was in fact the best strategy for Player A to maximize his earnings because the punishment amount decreased with greater transfers (with zero punishment above a transfer of 50 MUs) in an approximately linear fashion (Fig. 2).

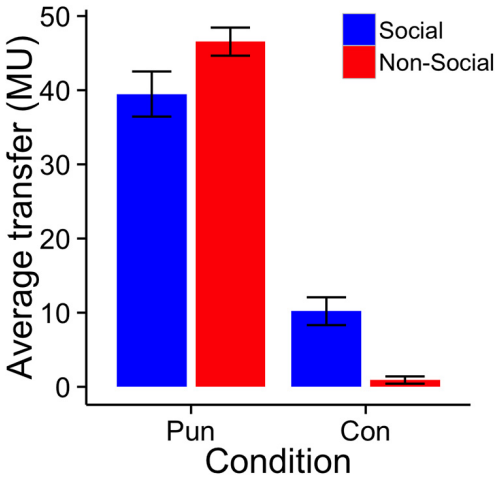

Figure 1 Amounts transferred by Player A in the PUN and the CON condition of both the social and the nonsocial treatment. Transfers are represented in experimental monetary units out of a given amount of 100 units. Error bars represent the standard error of the mean for the group mean. Paired sample Wilcoxon signed rank tests (social W = 276, p = 2.88e−5; nonsocial W = 231, p = 6.36e−5) showed significant differences between the PUN and CON transfer rates in each group.

In our initial neuroimaging analysis, we examined the degree to which vmPFC activity reflected value computations during monetary transfer decisions in both treatment types using a general linear model on BOLD signals. This analysis showed a positive association between kept amounts and vmPFC BOLD activity (Fig. 3*a*; $p < 0.05$ whole-brain corrected) across all participants. In addition

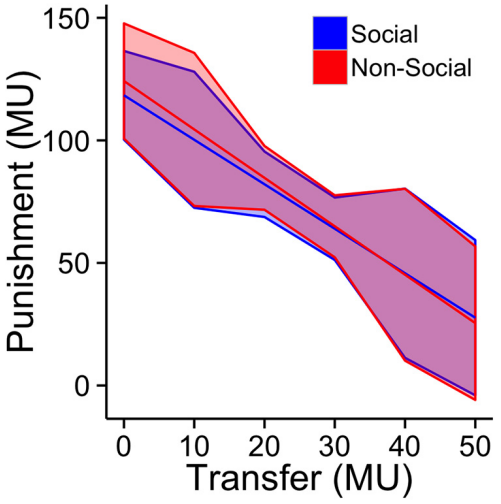

Figure 2 The plot shows punishment distributions as a function of amount transferred for both social (blue) and nonsocial groups (red). Punishment was regressed onto transfers up to 50 MUs, with the predicted punishment (thick line) and the SDs of the residuals (shaded area) for each transfer amount. Transfers >50 MU resulted in zero punishment. The overlapping distributions for the social and nonsocial treatments indicate that the computer algorithm was successful in replicating human punishment behavior.

to vmPFC, BOLD activity in dmPFC, right frontopolar cortex, and occipital regions also correlated with the amount kept at the time of choice (Table 1). The correlation between amount kept and BOLD activity in the vmPFC ROI was not significantly different between treatment groups (two-sample t test, $t_{(42)} = 1.4$, $p = 0.332$ uncorrected) indicating that participants playing against humans and computers represented the amount kept to an equal degree in vmPFC. Furthermore, there was no significant difference between the social and nonsocial groups in the correlation with amount kept and BOLD activity in any brain region after correcting for multiple comparisons.

The vmPFC result is consistent with theoretical models and existing empirical data suggesting a central role for vmPFC in the computation of subjective values for a wide range of decision contexts (Kable and Glimcher, 2009; Rangel and Hare, 2010; Rushworth et al., 2012; Bartra et al., 2013; Clithero and Rangel, 2013). Such theories also posit that if vmPFC acts as a general valuation system, then its interactions will be modulated such that coupling with regions providing decision relevant information will increase.

Next, we tested the hypothesis that the coupling between vmPFC and the right TPJ will increase more during decisions that require strategic evaluations of another person's response to the outcome than in complexity matched control conditions using a PPI analysis with the vmPFC as the seed region. This analysis examines whether the correlations between vmPFC activity and other brain regions differ in social versus nonsocial PUN transfer decisions. Note that in both the social and nonsocial PUN conditions participants need to make strategic transfer decisions that take into account Player B's likely level of punishment (i.e. fairness norm enforcement), and it is only the nature of Player B (human vs computer) that differs between groups. We found that participants in the social treatment showed more positive correlations between TPJ and vmPFC in PUN trials compared with the nonsocial treatment (Fig 3; $p < 0.05$ SVC; peak T = 3.97 at x, y, z = 60, −48, 16; extent = 115 voxels). *Post hoc* one-sample t tests showed that the average PPI effect in these voxels for social PUN was greater than zero ($t_{(20)} = 2.51$; $p = 0.021$), whereas the average PPI effect for nonsocial PUN was less than zero ($t_{(20)} = -3.79$; $p = 0.001$). Exploratory analyses revealed no other regions that showed this pattern of connectivity with vmPFC after correcting for multiple comparisons. However, for completeness, we also list regions exceeding a threshold determined by the lowest individual voxel t statistic ($t > 2.29$) derived from the right TPJ cluster (Table 2). Furthermore, there were no voxels that showed a significant PPI effect in either social or nonsocial CON trials after correcting for multiple comparisons within the independent TPJ ROI or in the entire volume.

To test whether vmPFC-TPJ PUN PPI strength is related to the overall strategic play of the participants, we tested whether the individual PPI difference contrast (PUN − CON) differentially correlated with participants' average punishment amounts in the social compared with nonso-

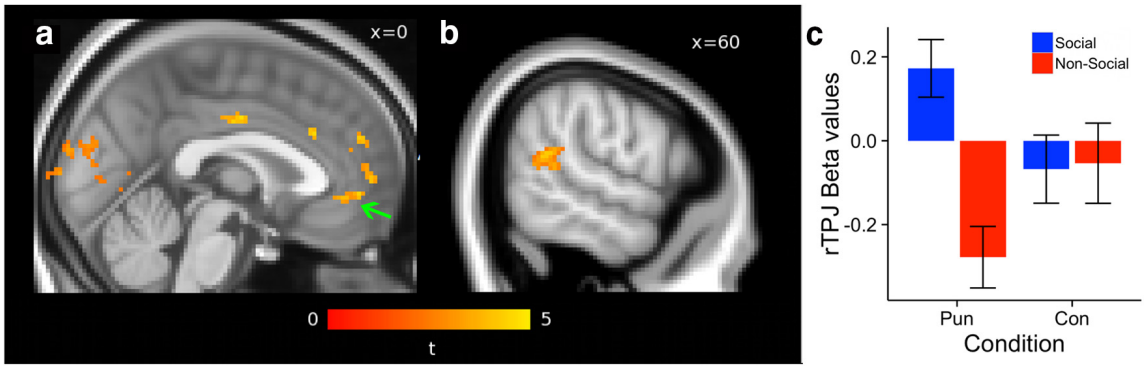

Figure 3 Activity and connectivity at the time of choice. ***a***, Regions showing a positive correlation with the amount of monetary units participants decided to keep on each trial across all decision types. The green arrow indicates the vmPFC cluster used to extract time courses for the PPI analysis. ***b***, Region of TPJ showing increased functional connectivity with vmPFC during strategic decisions made in social punishment compared with nonsocial punishment contexts. All voxels shown in ***a*** and ***b*** are significant at $p < 0.05$ after correcting for multiple comparisons. ***c***, Bar graph showing the relative coupling between vmPFC and TPJ by treatment group and choice context and demonstrating that increased TPJ-vmPFC coupling is specific to choices that are both strategic and social in nature. Error bars represent the standard error of the mean for the group mean. These bar plots are presented for visualization purposes only and were not used as a basis for any statistical analysis.

cial groups. This second level, between subjects regression analysis revealed a link between vmPFC-TPJ PPI during PUN trials and average punishment levels that were stronger in social more than nonsocial treatment participants. In the social group, greater vmPFC-TPJ PPI was associated with less punishment by Player B, whereas there was no significant relationship in the nonsocial group (Fig. 4; $p < 0.05$ SVC; peak T = 3.88 at $x, y, z = 56, -50, 16$; extent 8 voxels).

For completeness, we repeated our PPI analysis replacing the vmPFC seed with a region of dmPFC that also correlated with amount kept at the time of choice. We tested this dmPFC seed in addition to vmPFC because the dmPFC has been implicated in alternative value representation and strategic mentalizing processes (Frith and Frith, 2006; Hampton et al., 2008; Coricelli and Nagel, 2009; Nicolle et al., 2012; Zhu et al., 2012). However, we found no significant differences in connectivity with the dmPFC during social compared with nonsocial PUN trials within our TPJ ROI or at the whole-brain level after correcting for multiple comparisons.

We also examined brain activity at the time of outcome when subjects learned how much profit they had made in the previous trial. We found that the parametric regressor for profit magnitude at outcome (PR3 from GLM-1) correlated with BOLD activity in several regions, including bilateral striatum and left lateral frontal cortex ($p < 0.05$ whole-brain corrected; Table 3). Just as with the BOLD correlations at the time of choice, there were no regions showing a difference between the social and nonsocial groups in the correlation with profit magnitude at outcome. In addition, we conducted an ROI analysis on the BOLD correlation with profit at feedback using the voxels from the TPJ cluster showing the PUN PPI difference between the groups. We found that across all subjects there was a significantly negative effect of profit on TPJ activity at the time of feedback (one sample $t_{(41)} = -2.15$, $p = 0.037$), and once again, the groups did not significantly differ in this effect (two sample $t_{(40)} = 0.13$, $p = 0.90$).

Discussion

Our results indicate a role for vmPFC in the computation of value during strategic choices involving norm enforcement and suggest that increased TPJ-vmPFC coupling is especially important in decisions that involve strategic

Table 2 Location and extent of functional clusters showing a difference in PPI with vmPFC between social and nonsocial PUN decisions that was greater than or equal to the effect in our *a priori* TPJ region

Region	Hemisphere	Extent	*x*	*y*	*z*	Peak T
TPJ	R	144	60	−48	16	3.97
Parahippocampal gyrus	R	93	22	−26	−14	3.87
Lingual gyrus	R	86	28	−50	4	3.87
Fusiform cortex	L	78	−34	−38	−18	4.12
Fusiform cortex	L	66	−38	−2	−32	4.31
White matter/insular cortex	L	60	−28	−16	24	3.96
STG	L	51	−52	−24	6	3.36

Peak coordinates (*x*,*y*,*z*) are listed in MNI space. T values are test statistics derived from 5000 permutations of the data. Clusters reported are all of those that surpass a threshold set by lowest *t* value in the small volume corrected TPJ cluster ($t > 2.29$) and minimum cluster size of 50 voxels ($2 \times 2 \times 2$ mm). Note that these results are reported here for completeness only and are not corrected for multiple comparisons and thus not the subject of any inference in this paper. STG, Superior temporal gyrus.

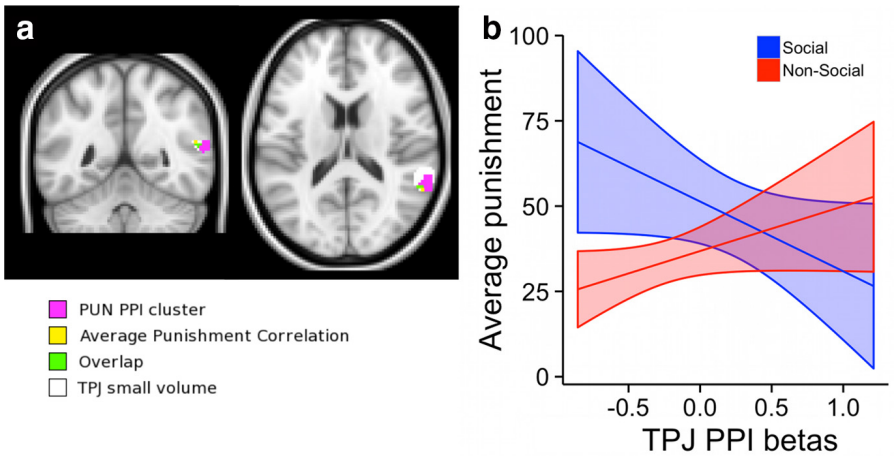

Figure 4 Regions of the TPJ relating to the vmPFC PPI at time of choice. ***a***, The voxels in magenta show greater increases in connectivity with vmPFC during PUN choices in the social compared with the nonsocial group, controlling for connectivity in CON trials. Voxels in yellow are those where the PUN − CON PPI was significantly correlated with individual average punishment levels in the social, but not the nonsocial group. Green voxels represent the overlap of both effects. Clusters are significant at $p < 0.05$ SVC within the TPJ small volume shown in white. ***b***, The fitted regression slopes between TPJ-vmPFC PPI at the time of choice and an individual's average punishment level separately for the social (blue) and nonsocial (red) groups. The shading around the regression lines indicates the 95% confidence intervals.

considerations of how social others will react to one's own actions. Despite the fact that participants were fully informed that the computer opponents were programmed to punish fairness norm violations at the same levels as real human players, the coupling between TPJ and vmPFC value computation regions did not increase in nonsocial PUN decisions, and in fact, decreased relative to the nondecision baseline.

This pattern of TPJ results is consistent with previous experiments showing that multivariate analyses of TPJ activity could be used to help predict bet and fold decisions in a simplified poker game against human opponents, but including TPJ activity measures actually decreased the model's predictive power for computer opponents (Carter et al., 2012). These previous experiments did not however examine the connectivity between TPJ and other brain regions. Our TPJ-vmPFC connectivity results demonstrate that in the realm of value-based choices, TPJ-vmPFC coupling increases during strategic choices when paired with human counterparts, but decreases with computer partners. Moreover, increased connectivity between vmPFC and TPJ at the time of choice is associated with more advantageous strategic decision-making (i.e. lower norm-enforcing punishment) in social but not nonsocial contexts. This is consistent with the idea that vmPFC incorporates information from distributed brain regions into value computations and that inputs are either enhanced or inhibited as a function of their relevance in the current state. Moreover, TPJ-vmPFC coupling did not significantly increase in either the social or the nonsocial CON trials where punishment predictions were not necessary because the opponent could not respond. This indicates that TPJ-vmPFC connectivity was not simply a function of the social nature of the decision, but rather occurred selectively when both social and strategic factors were in play.

Table 3 Regions correlating with profit at the time of feedback

Region	Hemisphere	Extent	*x*	*y*	*z*	Peak T
Insula/striatum	R/L	1645	32	12	4	6.97
Striatum	L	475	−30	−14	10	5.5
Frontopolar cortex	L	325	−38	60	4	4.88
Precentral gyrus	R	44	58	−4	22	4.17
Caudate tail	R	16	18	−4	26	3.85
Posterior insula	L	16	−38	−18	0	4.32
Parietal operculum	L	10	−48	−30	22	4.28

Peak coordinates (*x*,*y*,*z*) are listed in MNI space. T values are test statistics derived from 5000 permutations of the data. All regions are significant at $p < 0.05$ whole-brain familywise error corrected for multiple comparisons.

Previous work has shown that TPJ activity reflects social learning signals in the context of repeated interactions where it is advantageous to learn about other human players (Behrens et al., 2008; Hampton et al., 2008). This learning takes the form of update signals measuring deviations from the expected result at the time of feedback when decision outcomes are shown. These update or error signals are presumably used to guide subsequent choices when paired with the same person in the future, although the impact of TPJ activity at the time of subsequent choices was not explicitly examined in these previous reports. In the current paradigm, participants are paired with a different human partner on each trial, and therefore, outcomes of previous trial choices cannot be directly applied to future decisions. However, it may be that TPJ activity also plays a role in forming expectations based on average or normative behavior. There is a strong social norm for fairness and this norm could be used as a basis for predicting the degree of punishment by an unknown Player B that would result from various monetary splits. Consistent with this role, we found that TPJ activity increased when participants were shown feedback indi-

cating that a strategic adjustment was necessary (i.e. low profits) on the following choice to avoid future norm enforcing responses from Player B. Moreover, the results summarized in Figure 4 suggest that increased connectivity between TPJ and vmPFC may be a mechanism by which such predictions are incorporated into value computations at the time of choice.

In addition to vmPFC, BOLD activity in several other brain regions, particularly dmPFC, correlated with the amount kept for oneself when deciding how to allocate MUs on each trial. The correlation with kept amount in dmPFC is of particular interest given previous findings that activity in this region relates to individual differences in type or level of reasoning during social interactions (Hampton et al., 2008; Coricelli and Nagel, 2009; Zhu et al., 2012). Our findings in dmPFC build on these previous individual difference results and demonstrate that this region also reflects choice specific components of strategic valuation during decisions in which social norm compliance can be enforced through peer punishment. Although the current dataset was not designed to distinguish between the value related activity in regions such as vmPFC and dmPFC, previous reports have suggested that there is a dorsal to ventral gradient for modeled and executed value functions along the mPFC (Nicolle et al., 2012). If our subjects are engaging in predictive forecasting (i.e. modeling) of Player B's responses to their transfers and decisions are taken (i.e. executed) on the basis of these models, then this could explain why we find activity correlated with the amount kept in both ventral and dorsal portions of mPFC. However, further experiments will be necessary to test this speculative hypothesis.

One limitation of the current dataset is that there were a relatively small number of choices for each participant per condition ($n = 12$). Therefore, it is possible that future studies including more choices per participant, and thus having greater power, will find additional changes in vmPFC connectivity associated with social strategic decision-making.

Decisions that balance welfare for self with the impacts on and reactions of others to one's own choices are ubiquitous in social life. Our results provide insights into the neural mechanisms underlying such behavior and suggest a key role for interactions between TPJ and vmPFC. These findings are an important advance in our understanding of the neurobiology underlying strategic social choice and provide a basis for future investigations into this central aspect of human behavior.

References

Bartra O, McGuire JT, Kable JW (2013) The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage 76:412–427. CrossRef Medline

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. Nature 456:245–249. CrossRef Medline

Bhatt MA, Lohrenz T, Camerer CF, Montague PR (2010) Neural signatures of strategic types in a two-person bargaining game. Proc Natl Acad Sci U S A 107:19720–19725. CrossRef Medline

Camerer C (2003) Behavioral game theory: experiments in strategic interaction. Princeton, NJ: Princeton UP.

Carter RM, Huettel SA (2013) A nexus model of the temporal-parietal junction. Trends Cogn Sci 17:328–336. [PMC] [10.1016/j.tics.2013.05.007] [23790322]

Carter RM, Bowling DL, Reeck C, Huettel SA (2012) A distinct role of the temporal-parietal junction in predicting socially guided decisions. Science 337:109–111. CrossRef Medline

Clithero JA, Rangel A (2013) Informatic parcellation of the network involved in the computation of subjective value. Soc Cogn Affect Neurosci 9:1289–1302.

Coricelli G, Nagel R (2009) Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proc Natl Acad Sci U S A 106:9163–9168. CrossRef Medline

Frith CD, Frith U (2006) How we predict what other people are going to do. Brain Res 1079:36–46. CrossRef Medline

Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19:200–207. Medline

Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci U S A 105:6741–6746. CrossRef Medline

Hare TA, Camerer CF, Knoepfle DT, Rangel A (2010) Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J Neurosci 30:583–590. CrossRef Medline

Kable JW, Glimcher PW (2009) The neurobiology of decision: consensus and controversy. Neuron 63:733–745. CrossRef Medline

Krajbich I, Adolphs R, Tranel D, Denburg NL, Camerer CF (2009) Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. J Neurosci 29:2188–2192. CrossRef Medline

Morishima Y, Schunk D, Bruhin A, Ruff CC, Fehr E (2012) Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. Neuron 75:73–79. CrossRef Medline

Nicolle A, Klein-Flügge MC, Hunt LT, Vlaev I, Dolan RJ, Behrens TE (2012) An agent independent axis for executed and modeled choice in medial prefrontal cortex. Neuron 75:1114–1121. CrossRef Medline

Rangel A, Hare T (2010) Neural computations associated with goal-directed choice. Curr Opin Neurobiol 20:262–270. CrossRef Medline

Rilling JK, Sanfey AG (2011) The neuroscience of social decision-making. Annu Rev Psychol 62:23–48. CrossRef Medline

Rushworth MF, Kolling N, Sallet J, Mars RB (2012) Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Curr Opin Neurobiol 22:946–955. CrossRef

Saxe R, Wexler A (2005) Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia 43:1391–1399. CrossRef Medline

Smith SN, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing-threshold dependence and localisation in cluster inference. Neuroimage 44:83–98. CrossRef Medline

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289. CrossRef Medline

Zhu L, Mathewson KE, Hsu M (2012) Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. Proc Natl Acad Sci U S A 109:1419–1424. CrossRef Medline

Appendix B

Manuscript for study 2: Model based brain substrates of fairness and money

B.1 Introduction

Choices, especially those directly involving social factors, are influenced by numerous individual and context specific variables. Converging evidence from decision making studies across fields such as biology, economics, psychology, and neuroscience suggest that humans and other animals can integrate multiple attributes including rewards, punishments, costs, temporal delays, etc. into representations of the predicted value for each decision option using cognitive mechanisms and neural circuits that are at least partially overlapping across the domains of primary, secondary, and social goods (Chib et al. 2009; Basten et al. 2010; Hare et al. 2010; Bartra et al. 2013). The Ultimatum game (UG; Gth et al. 1982) is a widely used paradigm for studying social decisions in which monetary gain and fairness (i.e. equality) attributes both play important roles in determining players choices (see Methods; Fehr and Camerer 2007). The finding that people do not behave according to the Nash equilibrium (Mailath 1998) predicted by assumptions of purely selfish preferences in the UG is extremely robust (List 2007; Slonim and Roth 1998; Cameron 1999; Andersen et al. 2011). However, the precise proportion of the total amount proposers offer and responders are willing to accept varies across both individuals and choice contexts, indicating that expressed social preferences are malleable and at least partially state-dependent (Chang and Sanfey 2013; Wright et al. 2011; Sanfey 2009; Hoffman et al. 2000; Henrich et al. 2001; Andersen et al. 2011).

Previous UG experiments have shown that both recent experience and cognitive strategies can affect participants behavior and neural activity. For example, the same ultimatum is more likely to be accepted if presented following a string of less equitable offers than in the context of more equitable offers, and both objective and contextual aspects of fairness are reflected in the insula (Wright et al. 2011). The inferred intention behind the proposal is another type of context that affects responders accept/reject decisions (Falk et al. 2008). Using the explicit cognitive strategy of reappraising the intentions of the proposer as more negative or positive has been shown to recruit lateral prefrontal cortex (PFC) activity and to decrease and increase offer acceptance rates, respectively (Wout et al. 2010; Grecucci et al. 2013). Exogenously altering lateral PFC activity via transcranial magnetic stimulation (TMS) also changes the acceptance rate of unfair offers, despite leaving the fairness judgments of those offers unchanged (Knoch et al. 2006; Baumgartner et al. 2011). Even simple suggestions to consider the other players actions alter Ultimatum game behavior. When prompting proposers to consider re-

sponders expectations and subsequent responses, Hoffman and colleagues (2000) found a significant increase in offer magnitudes, while Anderson and colleagues (2011) were able to elicit reduced offer magnitudes from proposers using instructions that indicated rational responders should accept any non-zero offer. These findings suggest that choice contexts and attentional frames have a substantial impact on participants decision making processes during social interactions.

Recent work has utilized sequential sampling models (SSMs) to examine social decision making processes and provided novel insights into interpersonal behavior. Originally applied to perceptual and categorization tasks, sequential sampling models have been extended to the domain of value-based choices (Forstmann et al. 2016), and more recently social decision making. Thus far, SSMs have been shown to account for behavior in both two-person asset allocation decisions (e.g. UG) and charitable donation decisions (Krajbich et al. 2015; Hutcherson et al. 2015). In fact, it has been shown that the parameters of an SSM fit to decisions over primary rewards for self in one sample of participants can accurately predict the choices and reaction times of a separate group of participants playing in the role of responder in the UG (Krajbich et al. 2015). Critically, a parameter capturing the difference in the decision weights placed on visually fixated relative to non-fixated food items was assumed to apply to payoffs for oneself relative to the other player when making the out-of-sample predictions about UG behavior. The successful translation of a model capturing the influence of visual fixation, a proxy for attention, on food choice into the domain of interpersonal decision making suggests that attention or cognitive focus also plays an important role interpersonal social choices.

Here, we tested the hypothesis that simple cues directing the responders initial focus towards either the potential monetary gain or fairness of the outcome would alter decision making and its underlying neural activity during the Ultimatum game. We measured blood oxygen level dependent (BOLD) signals using functional magnetic resonance imaging (fMRI) while human participants made decisions about unfair ultimatums under choice conditions that directed participants initial focus towards offer magnitude or equality. Ultimatum acceptance rates increased in the money condition and decreased in the fairness focus condition relative to control trials. We estimated the parameters of an SSM to capture the changes in choice patterns and reaction times across the three conditions and then used these context-specific parameters to identify brain regions that represented decision signals in favor of accepting or rejecting an offer. Overall, a set of brain regions in-

cluding anterior cingulate cortex (aCC), dorsomedial frontal cortex (dmFC) and bilateral insula were similarly active during responders choices across the three conditions, but the degree to which BOLD signals were associated with monetary gains and fairness levels changed across choice contexts. These results provide further evidence for the brains ability to flexibly incorporate various outcome attributes (e.g. personal gain, equality) into integrated decision values as a function of environmental and individual states.

B.2 Materials and methods

B.2.1 Participants

In total, 34 healthy, right-handed adults performed our ultimatum game while undergoing fMRI scanning. Participants were screened for fMRI contraindications including acute medical conditions and psychiatric or neurological illness. All participants provided written informed consent in accordance with the human participants committee in Zurich. Ten participants were excluded from all behavioral and fMRI analyses because their choices (accept, reject) did not meet our *a priori* criterion for choice variation (each response must be chosen on at least 10 % of trials), leaving 24 participants for all analyses.

B.2.2 Behavioral paradigm

Participants played a modified Ultimatum Game. This game consisted of two players: player A who was endowed with a sum of money and was able to offer a portion of this endowment to player B. Player B was then given the option to accept the offer or reject it, causing the whole endowment to be returned to the experimenter. We used fMRI scanning while subjects playing the role of player B evaluated 54 different ultimatum offers. The modification to the original Ultimatum game consisted of an attentional focus manipulation. Every offer was repeated over three focus conditions (Money, Fairness and Natural), presented as blocks of 9 trials (162 trials in total; Figure B.1)). One trial was selected at random to be the one that counted for the participants payoff. The total endowment varied on each trial in order to minimize the correlation (r = 0.40) between monetary amount offered and the fairness

of the ultimatums. During the task, subjects had 3 seconds to accept or reject a given offer, and were informed that missed trials would be treated as rejections.

B.2.3 Behavioral analyses

We analyzed participants decision processes in each condition using a hierarchical Drift Diffusion Model that allowed us to model both reaction times and choice outcomes. To this end, we estimated a drift diffusion model for each condition, where the outcomes were accept or reject, and the starting bias and input function to the drift rate were allowed to vary. The input function was a linear combination of an intercept term, z-scored magnitude and fairness where the weights on each term were allowed to vary across conditions and subjects. The hierarchical estimation method used allowed us to consider each condition as a plate containing individual trial parameters, while conserving dependencies on other drift diffusion parameters assumed to be constant within a subject across conditions. Across subjects, the bias term and each of the weights on the input function to the drift rate was significantly greater than the null distribution. Similar to the logistic regression, there was a greater effect of offer magnitude in the money focus condition, and a greater effect of offer fairness in the fairness focus compared to the other conditions.

B.2.4 fMRI data acquisition

Images were acquired using a Philips Achieva 3T whole-body scanner with an eight-channel sensitivity-encoding head coil (Philips Medical Systems) at the Laboratory for Social and Neural Systems Research, University Hospital Zrich. Stimuli were presented to subjects using the MATLAB 2012b toolbox; Psychophysics Toolbox Software (Psychtoolbox 3.0, Brainard 1997) running on a stimulus PC that was connected to both a projector and a 932 interface & power supply (Current Design). This interface was also linked to a 4-button diamond fiber optic response pad and the scan computer to register scan pulses sent at the beginning of each volume acquisition. The paradigm was presented via a back-projection system mounted on the head coil. T2* weighted echo-planar images (EPIs) were acquired (34 slices per volume, Field of View 200 x 104.8 x 200 mm, slice thickness 2.5 mm, .6 mm gap, in-plane resolution 2.5*2.5 mm, matrix 80*80, repetition time 2000 ms, echo

time 30 ms, flip angle 77) and a SENSE reduction factor of 2. Volumes were acquired slice by slice in the axial orientation and ascending order at a +15 degree tilt to the line between the anterior and posterior commissures. In total, 390 volumes were collected over two experimental runs with five dummy volumes acquired at the start of each run to help mitigate the effect of scanner warm up. A T1 weighted turbo field echo structural image was acquired in sagittal orientation for each participant at the end of the scanning session (181 slices, Field of View 256 x 256 x 181 mm, slice thickness 1 mm, no gap, in-plane resolution 1*1 mm, matrix 256*256, repetition time 8.4 ms, echo time 3.89 ms, flip angle 8). B0/B1 maps of the phase and magnitude of the magnetic field were also derived from a fast field echo sequence (short echo time = 4.29 ms, long echo time = 7.4 ms) acquired prior to the first run of EPI volumes. Structural scans were coregistered to their mean EPIs and averaged together to permit anatomical localization of the functional activations at the group level. We measured breathing frequency and took an electrocardiogram with the in-built system of the scanner in order to correct for physiological noise.

B.2.5 fMRI analyses

The first GLM (GLM-D1) was computed in order to test for an overall effect of the drift rate. All trials were modelled as boxcars with onsets as the decision screen began and durations equal to the reaction time of that decision (missed trials were modeled with a duration of three seconds) convolvd with a canonical haemodynamic response function. The regressor of interest in GLM-D1 was the drift rate for all trials where a decision was made. In addition, dummy regressors were included for all trials as well as for rejected, money focus, fair focus and missed trials. These were included in order to account for potentially differential main effects associated with behaviour and conditions during the experiment. A dummy regressor for the instruction screens was also included with onsets at the start of the instruction screen with a duration of five seconds.

A second GLM (GLM-AC) was also computed to examine the effects of fairness and money in the attention conditions (AC) more closely. The dummy regressors were designed in a similar way to GLM-D1, but instead of a dummy regressor with all trials, one was included for the control trials alone. GLM-AC did not test drift rate regressors, but used the money and fairness presented in the decisions instead, split into separate regressors for

the attention conditions of money, fairness and control. Thus six parametric regressors were included in GLM-AC, three for the money correlation in each condition and three for the fairness correlation in each condition.

Finally, a second GLM (GLM-D2) was computed to further explore the relationship between the drift rate and BOLD activity. Since drift rate is related to accepting an offer, GLM-D2 split the drift rate regressor from GLM-D1 into accepted and rejected trials in order to test them separately. The dummy regressors were similar to GLM-D1, except for the dummy regressor for all trials being replaced with two dummy regressors for accepted and rejected trials.

Six motion regressors were added to each GLM before being computed using the SPM 8 software at the indivdual subject level. Group level non-parametric analysis was carried out using the FSL randomise function (Winkler et al. 2014) combining threshold free cluster enhancement (Smith and Nichols 2009) with 5000 permutations of the data to determine a null distribution for voxel-wise multiple comparisons correction and achieve statistical significance levels of $p < 0.05$ at the whole brain level.

B.3 Results

B.3.1 Behavioral

Participants choice behavior was similar to that reported in previous studies using the ultimatum game (C. F. Camerer 2003; Fehr and Camerer 2007). Note that all offers in our experiment were less than 50% of the total pot (i.e. unfair) and ranged in relative equality from 10 to 30% of the total pot. The mean acceptance rate for all offers in the baseline condition collapsing across both monetary gain and equality levels was 43.2% (95% highest density interval (HDI) = [35.1, 51.6]). Participants willingness to accept the unfair offers changed when explicitly focusing on money or equality before making their choices. When focusing on money, participants acceptance rates increased by 8.9% (95% HDI = [3.1, 14.3], posterior probability of an increase greater than 0 = 99.8%). In contrast, when focusing on equality, participants acceptance rates decreased by 4.2% (95% HDI = [-8.0, -0.5], posterior probability of a decrease greater than 0 = 98.7%).

We examined the influence of offer magnitudes and relative equality on choices and RTs in each condition using a Bayesian hierarchical drift diffusion modeling approach. The drift diffusion model (DDM) is a specific form of sequential sampling model that takes as input the relative evidence for each option and accumulates this evidence over time until reaching a decision boundary. Within our modeling framework, the relative evidence is computed as a function of the offer magnitude and equality on each trial. We initially compared two DDM models, one in which the coefficients weighting the influence of offer magnitude and equality on drift rate (i.e. relative evidence accumulation) and all other DDM parameters were constant for all trials and another in which the drift-rate coefficients and all other DDM parameters could vary in the three conditions, Money, Equality and Baseline. Model comparison based on the deviance information criterion (DIC) showed that the condition specific drift-rate model provided the best fit to the data (constant model DIC = 551690; condition specific model DIC = 346776; lower DIC indicates a better fit). The drift rate alone predicted subject responses 75% of the time and Figure B.2 shows the alignment between RT patterns predicted by our DDM and participants behavior. Comparisons of the coefficients reflecting the influence of offer magnitude and equality on choices in each condition revealed that offer magnitude had more impact during the Money focus condition relative to both neutral focus and Baseline trials, while the influence of equality was greater when the initial focus was directed towards the fairness attribute (Figure B.3). The drift-rate coefficients and bias differed significantly across choice conditions (tables B.1 and B.2).

B.3.2 fMRI

We first tested for associations between BOLD signals and the trial-specific drift rates estimated from choices and RTs at the behavioral level using the model labeled GLM-D1 in the Methods section. This model included a single onset for all trials regardless of decision or condition and a parametric regressor equal to the drift rate on each trial. This drift regressor identified regions that reflected an integration of each trials monetary and fairness attributes because the rate of evidence accumulation (drift) in our DDM varies as a function of the trial-specific offer equality levels and magnitudes as well as the choice context (Baseline, Fairness or Money conditions). Multiple brain regions exhibited a BOLD signal that correlated positively with the regressor for drift rate, including the amygdala, dlPFC, dorsal and ventral portions of

the medial PFC, fronto-polar cortex, insula (anterior and mid), precuneus, parietal cortex, thalamus, and ventral striatum (Figure B.4a; Table B.3). Note that within our modeling framework, the drift rate is proportional to the evidence in favor of accepting over rejecting the proposed monetary split.

A specification of the drift-diffusion model allowing for condition specific changes in the influence of offer magnitudes and fairness provided the best fit to the participants pattern of accept and reject decisions. Thus, variation in the effects of magnitude and fairness on choice is captured by the parametric regressor for drift-rate. To visualize the differential associations between BOLD activity and offer magnitude or fairness as a function of the attentional focus condition, we extracted the regression coefficients from a another model (GLM-AC) in selected functional ROIs identified by GLM-D1. GLM-AC included separate onsets for trials in each attention condition (AC) and parameteric regressors for the offer magnitude and fairness on each trial. To select voxels correlated with the drift rate regressor from GLM-D2 in an unbiased fashion, we split our data into two halves and computed the group-level contrast for drift rate in accept versus reject decisions in each half separately. We then extracted the regression coefficients representing the associations between offer magnitude or fairness and BOLD activity for each half of the sample based on the ROIs identified in the other half (see Method for further details). Figure B.4b shows the differential representations of offer magnitude and fairness in regions dmFC, frontal polar cortex and vmPFC as a function of the attention condition.

Similarly to previous literature, we tested areas of the brain exhibiting differenctial BOLD activity when accepting and rejecting in GLM-D2. Consistent with previous neuroimaging and transcranial stimulation studies, we found that accepting the unfair offers that were presented was associated with increased activity in several brain regions (Figure B.5a; Table B.4) including the right dorsolateral prefrontal cortex (dlPFC), anterior insula, supramarginal gyrus and striatum (putamen and caudate). On the other hand, the left temporal parietal junction and superior temporal sulcus showed greater BOLD signal when participants decided to reject the offer (Figure B.5b; table B.5).

Finally, using the same model (GLM-D2), we tested for brain regions that differentially represented the evidence for accepting the ultimatum as a function of the choice the participant ultimately made using a second model (GLM-D2) that split the trials into separate regressors according to participants accept and reject decisions. The contrast for drift rate in Accept

greater than Reject trials showed significant differences in regions such as vmPFC, striatum, and posterior cingulate cortex (see Table B.6) that have been shown to correlate with stimulus values for a wide range of goods and experiences (Bartra et al. 2013; Clithero and Rangel 2014) and to adapt value representations for multi-attribute stimuli according to context or attentional cues (Nicolle et al. 2012; Hare et al. 2011a; Rudorf and Hare 2014). Separate examinations of the Accept and Reject trials alone revealed that many regions displayed a negative association with drift rate (i.e. the relative evidence to accept the proposal) during trials in which participants chose to reject the offer (Figure B.6; Table B.7). This negative relationship between the drift rate and BOLD signals suggests that activity in these regions may reflect the evidence in favor of rejecting the offer.

B.4 Discussion

In this study, we used a modified ultimatum game in order to separate factors relating to fairness and monetary concerns. The primary aim was to examine the decision making process that balances these factors and potential neural mechanisms that support it. To this end, a DDM framework was applied to a two alternative choice ultimatum game. To test whether this model was sensitive to intra-individual weights on the fairness and monetary concerns, a simple framing manipulation was used. The resulting drift rate per decision led to an integrated measure of the decision variable which was tested against the BOLD data at the time of decision making.

Central to the question of how social decisions are made, is how different options can be compared when their attributes seem radically different, e.g. how much money is spending time with family worth? In the ultimatum game, the question is reduced to weighing financial gain against fairness (note that since we used only unfair offers, this is proportional to inequality). The DDM allows us to fit a single process to empirical data with quantitatively changing monetary gain and fairness. Using a hierarchical estimation of DDM parameters allowed us to not only make trial level fits of model parameters, but also to pool evidence for different framing conditions. Thus different parameters were fit to baseline, money and fairness conditions. Testing the model parameters over these conditions revealed that several aspects of the DDM was sensitive to these effects, including the bias and evidence required to accept an offer. The input function to the drift rate also revealed greater weights on the monetary gain factor when in the money frame and a

greater weight on the fairness factor when in the fairness frame.

These results demonstrate that attentional cues can influence social decision making. and suggest that this effect can be captured by a DDM. This effect was expected given previous results that have changed the framing of objectively similar offers (Wright et al. 2011). Several studies have demonstrated that the perception of fairness can be decoupled from its effects on behaviour. This appears to be true whether the manipulation is based on framing (Wright et al. 2011), altering neurotransmitter levels (Crockett et al. 2008) or directly interfering with neural activity (Knoch et al. 2006). This suggests that there may be a difference between a more abstract knowledge of fairness and how it is evaluated and integrated into social decision making. In this study we observed the latter effect.

Using the DDM fit, the trial by trial information was captured in the changing drift rate. This was used to correlate the BOLD activity at the time of these decisions. By testing the correlation of BOLD with drift rate when accepting, minus the drift rate when rejecting, it appears that the BOLD signal increases strongly in relation to the decision variable driving towards the decision that will be made. Seeing this throughout the medial frontal cortex ties in with other research on regions of the brain that relate to the value integration of a choice (Hare et al. 2011a; Nicolle et al. 2012; Rudorf and Hare 2014). Together with the BOLD correlations with drift rate when rejecting the unfair offers presented in this study, it appears that ultimatum game rejections may be rational and consistent when including fairness concerns.

While the DDM has been previously applied to value based decision making in a social context (Krajbich et al. 2015), our study shows how this mechanism may be supported by neural processes. This is of particular interest from an economic perspective, as it adds weight to the argument that economic decisions should be described with process models in order to better understand how and when choices will be sensitive to specific contextual factors. Building on this research, it may be possible to develop parsimonious models to describe the process of satisficing over multiple domains, helping to solve problems involving bounded rationality(Munier et al. 1999).

B.5 References

Andersen, Steffen et al. (2011). "Stakes Matter in Ultimatum Games". In: *The American Economic Review* 101.7, pp. 3427–3439. ISSN: 0002-8282. URL: http://www.jstor.org/stable/41408744.

Bartra, Oscar et al. (Aug. 2013). "The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value". In: *NeuroImage* 76, pp. 412–427. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.02.063. URL: http://www.sciencedirect.com/science/article/pii/S1053811913002188.

Basten, Ulrike et al. (Dec. 2010). "How the brain integrates costs and benefits during decision making". en. In: *Proceedings of the National Academy of Sciences* 107.50, pp. 21767–21772. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0908104107. URL: http://www.pnas.org/content/107/50/21767 (visited on 08/24/2017).

Baumgartner, Thomas et al. (Nov. 2011). "Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice". en. In: *Nature Neuroscience* 14.11, pp. 1468–1474. ISSN: 1097-6256. DOI: 10.1038/nn.2933. URL: http://www.nature.com/neuro/journal/v14/n11/full/nn.2933.html (visited on 07/28/2017).

Brainard, D. H. (1997). "The Psychophysics Toolbox". eng. In: *Spatial Vision* 10.4, pp. 433–436. ISSN: 0169-1015.

Camerer, Colin F. (2003). "Ultimatum and Dictator Games: Basic Results". In: *Behavioral Game Theory: Experiments in Strategic Interaction*, pp. 48–58.

Cameron, Lisa A. (Jan. 1999). "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia". en. In: *Economic Inquiry* 37.1, pp. 47–59. ISSN: 1465-7295. DOI: 10.1111/j.1465-7295.1999.tb01415.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1465-7295.1999.tb01415.x/abstract.

Chang, Luke J. and Alan G. Sanfey (Mar. 2013). "Great expectations: neural computations underlying the use of social norms in decision-making". In: *Social Cognitive and Affective Neuroscience* 8.3, pp. 277–284. ISSN: 1749-5016. DOI: 10.1093/scan/nsr094. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3594719/.

Chib, Vikram S. et al. (Sept. 2009). "Evidence for a Common Representation of Decision Values for Dissimilar Goods in Human Ventromedial Prefrontal Cortex". en. In: *Journal of Neuroscience* 29.39, pp. 12315–12320. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2575-09.2009. URL: http://www.jneurosci.org/content/29/39/12315 (visited on 08/24/2017).

Clithero, John A. and Antonio Rangel (Sept. 2014). "Informatic parcellation of the network involved in the computation of subjective value". In: *Social Cognitive and Affective Neuroscience* 9.9, pp. 1289–1302. ISSN: 1749-5016. DOI: 10.1093/scan/nst106. URL: https://academic.oup.com/scan/article/9/9/1289/1675099/Informatic-parcellation-of-the-network-involved-in (visited on 08/05/2017).

Crockett, Molly J. et al. (June 2008). "Serotonin modulates behavioral reactions to unfairness". In: *Science (New York, N.Y.)* 320.5884, p. 1739. ISSN: 0036-8075. DOI: 10.1126/science.1155577. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2504725/.

Falk, Armin et al. (Jan. 2008). "Testing theories of fairnessIntentions matter". In: *Games and Economic Behavior* 62.1, pp. 287–303. ISSN: 0899-8256. DOI: 10.1016/j.geb.2007.06.001. URL: http://www.sciencedirect.com/science/article/pii/S0899825607000784.

Fehr, Ernst and Colin F. Camerer (Oct. 2007). "Social neuroeconomics: the neural circuitry of social preferences". In: *Trends in Cognitive Sciences* 11.10, pp. 419–427. ISSN: 1364-6613. DOI: 10.1016/j.tics.2007.09.002. URL: http://www.sciencedirect.com/science/article/pii/S136466130700215X.

Forstmann, B.U. et al. (2016). "Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions". In: *Annual review of psychology* 67, pp. 641–666. ISSN: 0066-4308. DOI: 10.1146/annurev-psych-122414-033645. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5112760/.

Grecucci, Alessandro et al. (Feb. 2013). "Reappraising the Ultimatum: an fMRI Study of Emotion Regulation and Decision Making". In: *Cerebral Cortex* 23.2, pp. 399–410. ISSN: 1047-3211. DOI: 10.1093/cercor/bhs028. URL: https://academic.oup.com/cercor/article/23/2/399/285843/Reappraising-the-Ultimatum-an-fMRI-Study-of (visited on 06/23/2017).

Gth, Werner et al. (Dec. 1982). "An experimental analysis of ultimatum bargaining". In: *Journal of Economic Behavior & Organization* 3.4, pp. 367–388. ISSN: 0167-2681. DOI: 10.1016/0167-2681(82)90011-7. URL: http://www.sciencedirect.com/science/article/pii/0167268182900117.

Hare, Todd A. et al. (Jan. 2010). "Value Computations in Ventral Medial Prefrontal Cortex during Charitable Decision Making Incorporate Input from Regions Involved in Social Cognition". en. In: *Journal of Neuroscience* 30.2, pp. 583–590. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4089-09.2010. URL: http://www.jneurosci.org/content/30/2/583 (visited on 06/23/2017).

Hare, Todd A. et al. (July 2011a). "Focusing Attention on the Health Aspects of Foods Changes Value Signals in vmPFC and Improves Dietary Choice". en. In: *Journal of Neuroscience* 31.30, pp. 11077–11087. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.6383-10.2011. URL: http://www.jneurosci.org/content/31/30/11077 (visited on 08/05/2017).

Henrich, Joseph et al. (2001). "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies". In: *The American Economic Review* 91.2, pp. 73–78. ISSN: 0002-8282. URL: http://www.jstor.org/stable/2677736.

Hoffman, Elizabeth et al. (June 2000). "The Impact of Exchange Context on the Activation of Equity in Ultimatum Games". en. In: *Experimental Economics* 3.1, pp. 5–9. ISSN: 1386-4157, 1573-6938. DOI: 10.1023/A:1009925123187. URL: https://link.springer.com/article/10.1023/A:1009925123187 (visited on 07/28/2017).

Hutcherson, Cendri A. et al. (July 2015). "A Neurocomputational Model of Altruistic Choice and Its Implications". English. In: *Neuron* 87.2, pp. 451–462. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2015.06.031. URL: http://www.cell.com/neuron/abstract/S0896-6273(15)00594-2 (visited on 07/28/2017).

Knoch, Daria et al. (Nov. 2006). "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex". en. In: *Science* 314.5800, pp. 829–832. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1129156. URL: http://science.sciencemag.org/content/314/5800/829 (visited on 07/28/2017).

Krajbich, Ian et al. (Oct. 2015). "A Common Mechanism Underlying Food Choice and Social Decisions". In: *PLOS Computational Biology* 11.10, e1004371. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004371. URL: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004371 (visited on 07/28/2017).

List, JohnA. (June 2007). "On the Interpretation of Giving in Dictator Games". In: *Journal of Political Economy* 115.3, pp. 482–493. ISSN: 0022-3808. DOI: 10.1086/519249. URL: http://www.journals.uchicago.edu/doi/abs/10.1086/519249.

Mailath, George (1998). "Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory". In: *Journal of Economic Literature* 36.3, pp. 1347–1374. URL: http://econpapers.repec.org/article/aeajeclit/v_3a36_3ay_3a1998_3ai_3a3_3ap_3a1347-1374.htm.

Munier, Bertrand et al. (Aug. 1999). "Bounded Rationality Modeling". en. In: *Marketing Letters* 10.3, pp. 233–248. ISSN: 0923-0645, 1573-059X. DOI: 10.1023/A:1008058417088. URL: https://link.springer.com/article/10.1023/A:1008058417088 (visited on 08/25/2017).

Nicolle, Antoinette et al. (Sept. 2012). "An Agent Independent Axis for Executed and Modeled Choice in Medial Prefrontal Cortex". English. In: *Neuron* 75.6, pp. 1114–1121. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2012.07.023. URL: http://www.cell.com/neuron/abstract/S0896-6273(12)00674-5 (visited on 08/05/2017).

Rudorf, Sarah and Todd A. Hare (Nov. 2014). "Interactions between Dorsolateral and Ventromedial Prefrontal Cortex Underlie Context-Dependent Stimulus Valuation in Goal-Directed Choice". en. In: *Journal of Neuroscience* 34.48, pp. 15988–15996. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3192-14.2014. URL: http://www.jneurosci.org/content/34/48/15988 (visited on 08/05/2017).

Sanfey, Alan G. (June 2009). "Expectations and social decision-making: biasing effects of prior knowledge on Ultimatum responses". en. In: *Mind & Society* 8.1, pp. 93–107. ISSN: 1593-7879, 1860-1839. DOI: 10.1007/s11299-009-0053-6. URL: https://link.springer.com/article/10.1007/s11299-009-0053-6 (visited on 07/28/2017).

Slonim, Robert and Alvin E. Roth (1998). "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic". In: *Econometrica* 66.3, pp. 569–596. ISSN: 0012-9682. DOI: 10.2307/2998575. URL: http://www.jstor.org/stable/2998575.

Smith, Stephen M. and Thomas E. Nichols (Jan. 2009). "Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference". eng. In: *NeuroImage* 44.1, pp. 83–98. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2008.03.061.

Winkler, Anderson M. et al. (May 2014). "Permutation inference for the general linear model". In: *NeuroImage* 92, pp. 381–397. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2014.01.060. URL: http://www.sciencedirect.com/science/article/pii/S1053811914000913.

Wout, Mascha van t et al. (Dec. 2010). "The influence of emotion regulation on social interactive decision-making". In: *Emotion (Washington, D.C.)* 10.6, pp. 815–821. ISSN: 1528-3542. DOI: 10.1037/a0020069. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3057682/.

Wright, Nicholas D et al. (Apr. 2011). "Neural segregation of objective and contextual aspects of fairness". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.14, pp. 5244–5252. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.3138-10.2011. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3109551/.

B.6 Tables

Table B.1: Differential DDM parameter coefficients across conditions

Drift Coefficient	Condition	t value	p value
Offer	money>fairness	13.92	1.08×10^{-12}
	money>baseline	12.38	1.188×10^{-11}
Percentage offer	fairness>money	7.59	1.05×10^{-7}
	fairness>baseline	4.73	9.15×10^{-5}

Table B.2: ANOVA for DDM parameters over condition factors

Drift Parameter	F value	p value
alpha	4.28	0.0177
beta	5.74	0.00495
theta	0.344	0.710
intercept	6.78	0.002054
offer	113	2.2×10^{-16}
fairness	28.7	8.34×10^{-10}

Table B.3: Regions correlating with drift rate across the whole brain

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	t
1930	Accumbens	R	8	13	-3.5	7.19
	Middle Frontal Gyrus	R	48	33	21.3	4.22
825	Cingulate Gyrus	L/R	10.5	43	12	4.93
	Paracingulate Gyrus	L/R	3	10.5	55.4	4.17
252	Frontal Orbital Cortex	L	-29.5	28	-0.4	7.10
251	Occipital Cortex	L	-29.5	-59.5	49.2	4.63
117	Lingual Gyrus	L/R	3	-62	2.7	4.90
113	Thalamus	L/R	3	-14.5	-6.6	6.47
33	Precuneous Cortex	L	-17	-69.5	39.9	4.61
13	Occipital Cortex	R	38	-69.5	33.7	4.90
12	Precuneous Cortex	L/R	-2	-72	52.3	3.35

Peak coordinates (x,y,z) are listed in MNI space. T values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at $p < 0.05$ whole brain family wise error corrected for multiple comparisons

Table B.4: Regions with increased BOLD activity when accepting than when rejecting across the whole brain

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	t
194	Supramarginal Gyrus	R	48	-32	46.1	6.48
153	Occipital Cortex	L	-22	-62	36.8	6.19
107	Putamen	R	20.5	8	5.8	5.57
102	Middle Frontal Gyrus	R	43	33	27.5	4.54
52	Precentral Gyrus	R	53	13	30.6	4.54
21	Occipital Cortex	R	28	-62	55.4	4.98

Peak coordinates (x,y,z) are listed in MNI space. T values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at p <0.05 whole brain family wise error corrected for multiple comparisons

Table B.5: Regions with increased BOLD activity when rejecting than when accepting across the whole brain

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	t
63	Angular Gyrus	L	-42	-52	27.5	5.80
41	Middle Temporal Gyrus	L	-59.5	-39.5	2.7	5.54

Peak coordinates (x,y,z) are listed in MNI space. T values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at p <0.05 whole brain family wise error corrected for multiple comparisons

Table B.6: Regions correlating with drift rate when accepting - drift rate when rejecting across the whole brain

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	t
18380	Angular Gyrus	R	50.5	-59.5	33.7	5.50
	Occipital Cortex	L	-39.5	-64.5	27.5	4.75
	Posterior Cingulate Gyrus	L/R	3	-37	30.6	4.72
	Fusiform Cortex	R	43	-49.5	-19	4.82
	Occipital Pole	R	10.5	-89.5	24.4	3.83
	Frontal Pole	R	18	68	5.8	5.39
	Middle Temporal Gyrus	L	-64.5	-37	-9.7	4.94
	Middle Frontal Gyrus	R	30.5	25.5	49.2	6.31
	Frontal Medial Cortex	L	-9.5	35.5	-22.1	4.85
	Frontal Pole	L	-47	48	8.9	4.49
	Amygdala	R	20.5	-4.5	-19	5.57
	Superior Frontal Gyrus	L	-24.5	20.5	46.1	4.92
	Inferior Frontal Gyrus	R	58	30.5	-3.5	3.86
70	Middle Temporal Gyrus	R	60.5	0.5	-25.2	4.83

Peak coordinates (x,y,z) are listed in MNI space. T values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at $p < 0.05$ whole brain family wise error corrected for multiple comparisons

Table B.7: Regions negatively correlating with drift rate when rejecting across the whole brain

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	t
1748	Occipital Cortex	L	48	-67	33.7	6.14
	Posterior Cingulate Gyrus	L/R	-4.5	-47	33.7	5.13
	Occipital Cortex	L	-37	-87	15.1	3.67
733	Frontal Pole	L	-7	58	30.6	4.79
	Frontal Medial Cortex	L	-7	38	-22.1	5.43
335	Postcentral Gyrus	R	65.5	-2	30.6	4.75
332	Occipital Cortex	L	-49.5	-67	27.5	6.22
44	Frontal Pole	L/R	13	70.5	12	4.71
34	Frontal Pole	R	35.5	38	-6.6	5.68
33	Middle Temporal Gyrus	R	68	-37	-3.5	4.16
21	Subcallosal Cortex	L/R	3	15.5	-9.7	4.49
13	Fusiform	R	33	-29.5	-22.1	4.85
11	Frontal Pole	L	-39.5	53	-0.4	3.67

Peak coordinates (x,y,z) are listed in MNI space. T values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at $p < 0.05$ whole brain family wise error corrected for multiple comparisons

B.7 Figure Legends

Figure B.1: Subjects completed a mini-block design task. The three conditions of the mini-blocks (a, b and c) are shown with the instruction, fixation and decision screens. The fixation and decision screens were repeated nine times per mini-block.

Figure B.2: Correlation ($r = 0.44$) between recorded reaction times and simulated reaction times based on the DDM model.

Figure B.3: boundary, non-decision time, bias and input function weights for the DDM fits for each condition, across subjects.

Figure B.4: Panel **a)** shows BOLD activity relating to the drift rate from GLM-D1 (across all trials). Panel **b)** uses ROIs from the analyis in a) to examine the drift rate in money, fairness and baseline conditions from GLM-AC. relate to the with BOLD activation positively correlating with drift rate across all trials. All voxels are shown at $p < 0.05$ familywise error corrected. Image shown in radiological space at MNI coordinates $(-4, 23, -1)$.

Figure B.5: Panels (a) and (b) show brain regions where BOLD activity was differentially correlated with the drift rate during trials in which the offer was accepted versus rejected. The warm color scale in **a)**represents voxels where the correlation with the trial-wise drift rate is more positive in accept relative to reject trials. The cool color scale in **b)** represents voxels where the correlation with the trial-wise drift rate is stronger in reject trials relative to accept trials. All voxels are shown at $p < 0.05$ familywise error corrected. The brain images are shown in radiological space at MNI coordinates $(8, 10, -1)$

Figure B.6: BOLD activation negatively correlating with drift rate in trials where the offer was rejected. All voxels are shown at $p < 0.05$ familywise error corrected. Image shown in radiological space at MNI coordinates $(1, -56, 6)$.

B.8 Figures

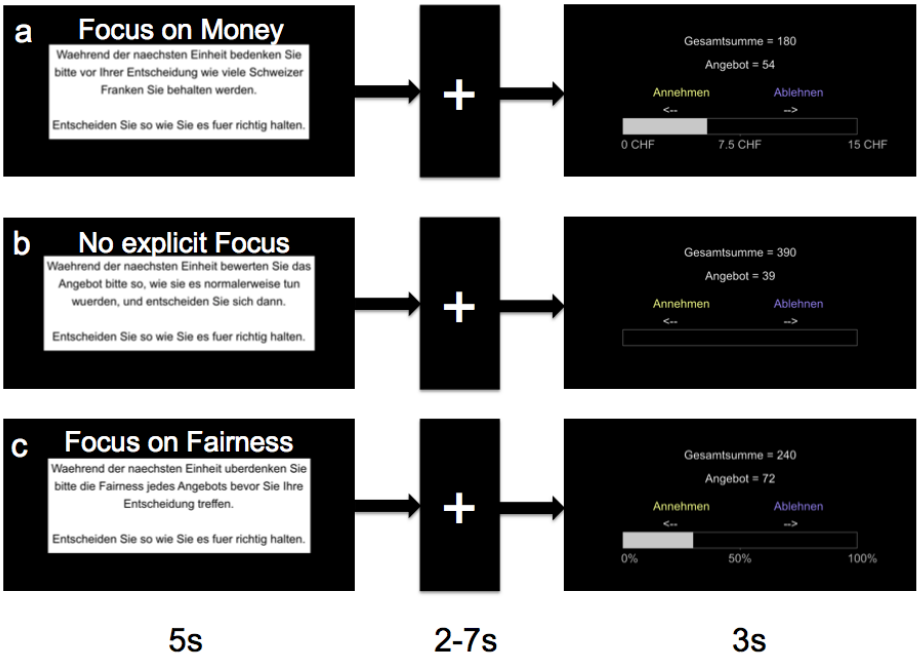

Figure B.1

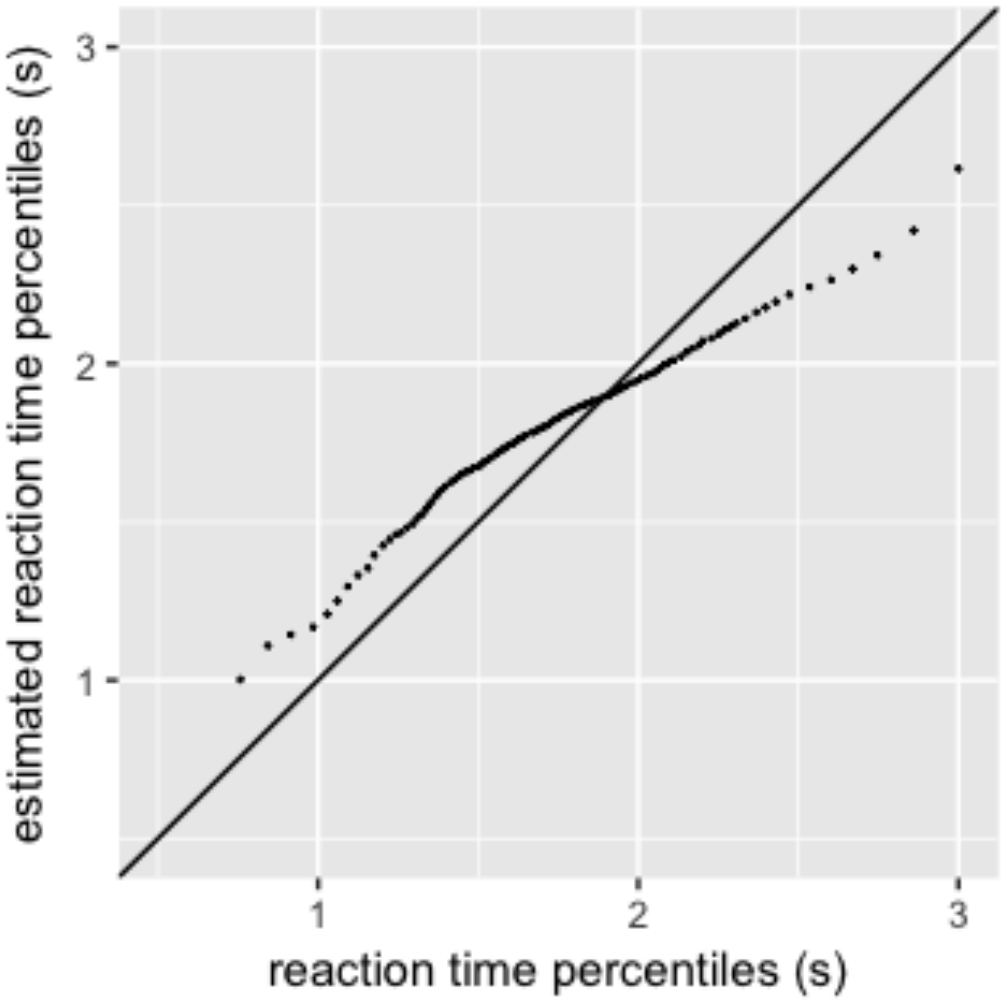
estimated reaction time percentiles (s)
3
2
1
1
2
3
reaction time percentiles (s)

Figure B.2

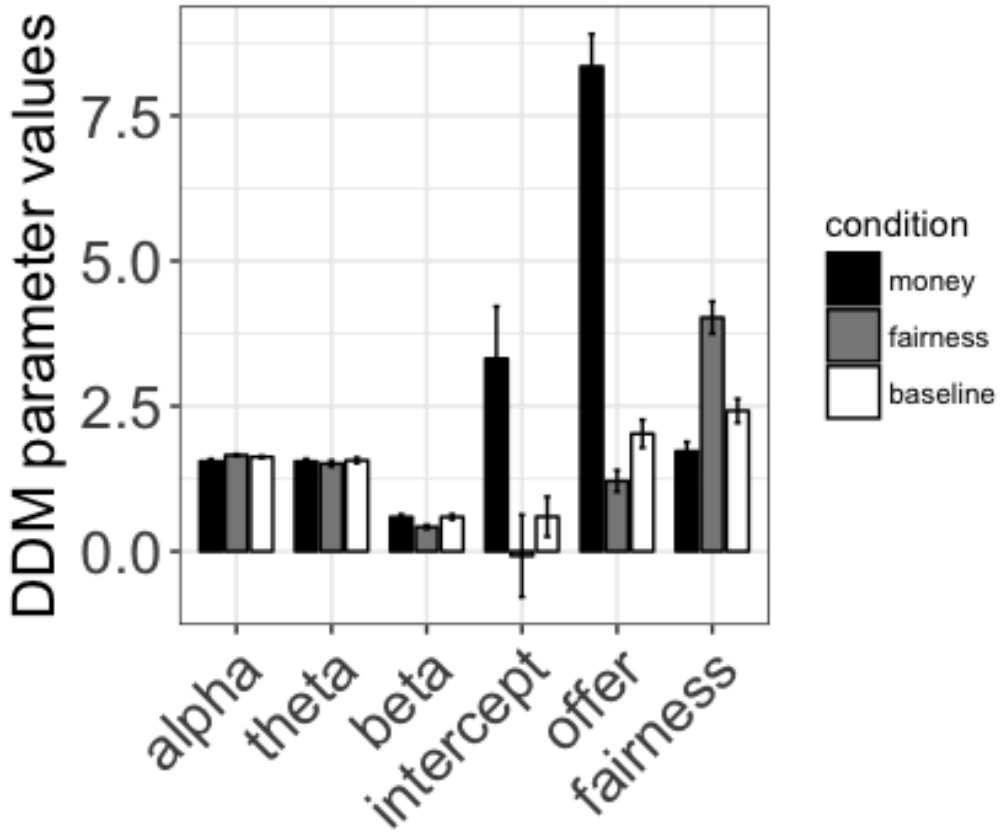

Figure B.3

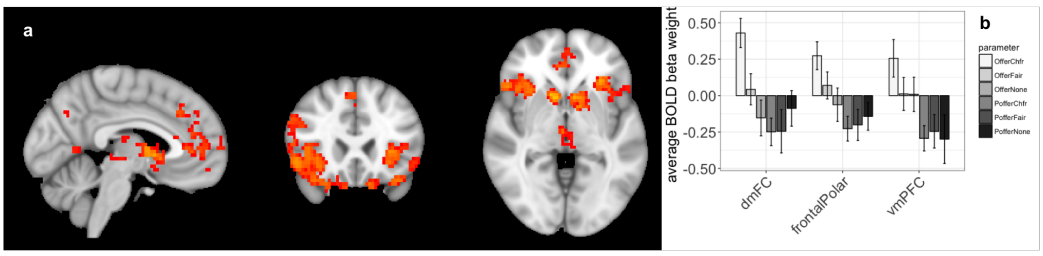

Figure B.4

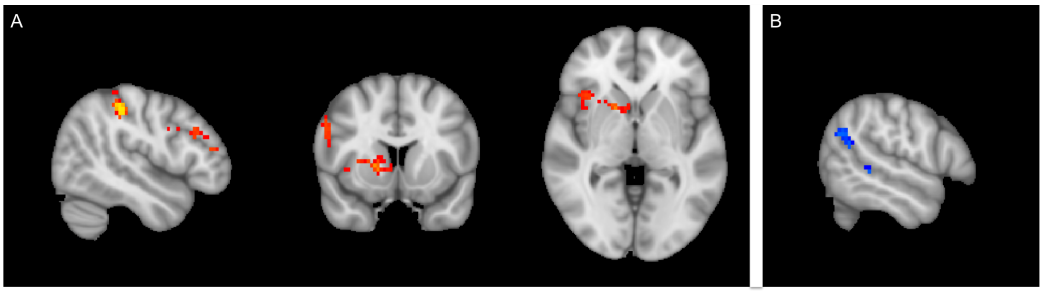

Figure B.5

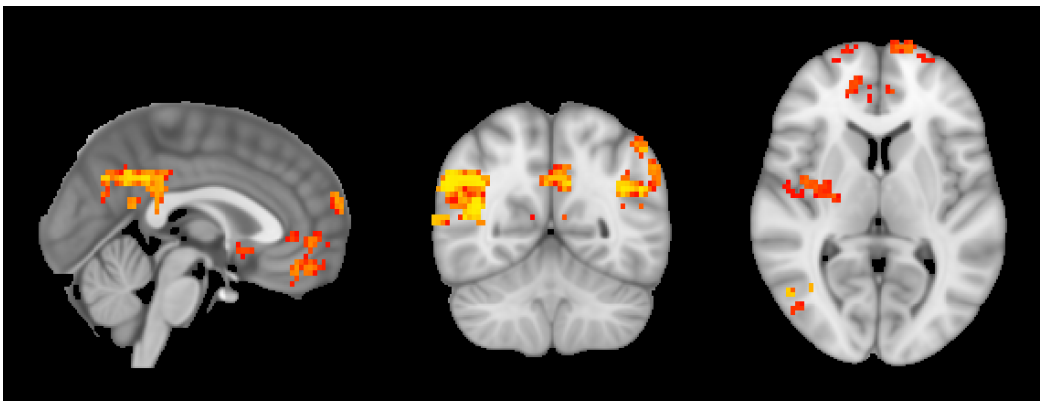

Figure B.6

Appendix C

Manuscript for study 3: Multivariate classification of decision features between charitable and personal decisions

C.1 Introduction

A key feature of human behaviour is our capacity for altruism. This describes a range of motivations and behaviours that lead to a net benefit to other individuals from our own actions. although altruism was defined over a century ago (Beesly 1875), it has become an active subject of study in psychology (Berkowitz 1972; Schwartz 1977; Batson and Moran 1999) and economics (Simon 1993; Andreoni 1990; Glazer and Konrad 1996) in the last few decades. Much work has been done on the taxonomy of altruistic behaviour in the psychological literature (see Feigin et al. 2014, for a systematic review), differentiating principally between pure altruism (no net benefit ot the self) and impure altruism (a net gain to the self as well as the other). These two types of altruism are then subdivided into several explicatory factors such as the role of emotion or expectations of reciprocity. The econonomic literature has taken a broadly similar path (Kagel and Roth 2016), but with more of an emphasis creating formal models to describe and explain the behaviour (Hubbard et al. 2016; Gino et al. 2016). In this paper, we restrict ourselves to one of the most prevalent forms of altruism, but also one of the most difficult to explain: charitable donation. This is often considered to be a pure altruism because the reputation and reciprocity effects can be controlled for, although it is possible that a 'warm glow' from charitable giving may deliver a primary reward to the subjects.

The primary goal of this paper is to better understand the low-level mechanisms that support these charitable decisions. This is an important step in understanding how altruistic concerns interact with other forms of value. Recent evidence suggests that interactions with wealth, goods and effort have different effects on the altruistic motive (Holmes et al. 2002; Strahilevitz 1999; Mayo and Tinsley 2009). One approach to shed light on the mechanisms underlying these interactions is to apply methods from neuroeconomics. Mounting evidence suggests that areas of the brain such as frontal cortex and striatum are vital to making value based decisions (Hare et al. 2010; Clithero and Rangel 2014). Several studies have used these methods in order to test whether there may be biological substrates of aspects of economic games such as the dictator game where there is an option to give money to another player who otherwise would not receive anything or charitable donations. For example (Izuma et al. 2009) found that being observed while deciding whether to donate to charity or keep the money for oneself increased charitable giving for difficult choices and that donating while under observation lead to a greater activation of the ventral striatum.

In this study, we present work where subjects could either keep an endowment, or donate to charities to or products they could buy in a task similar to the SHOP task (Knutson et al. 2007). This decision is considered as a tradeoff between the amount that the option is worth (willingness to pay, WTP), and the price offered for the option. Thus, this simplistic model of a value decision can be explored in two conditions - donating and buying, and the biological substrates of the attributes of the decision can be tested in either case. Here we present a functional magnetic resonance imaging (fMRI) study where the blood oxygenation level-dependent (BOLD) can be observed and tested against the experimental variables. Since there are multiple attributes that might need to be integrated in one region for a decision to take place, it has been proposed that local multivariate methods may aid in modeling these effects (Kahnt et al. 2011). Thus, in this study we apply a recently developed method, cross validated multivariate analysis of variance (cvMANOVA, Allefeld and Haynes 2014) to this data.

cvMANOVA is the multivariate (i.e. multiple dependant variables) extension of a standard analysis of variance, which subsequently applies a cross validation approach to reduce the extent of bias in the test statistic. This test statistic may be treated similarly to the univariate GLM test statistics commonly found in fMRI analyses. When applied to a fMRI searchlight analysis (Kriegeskorte et al. 2006), a sphere of BOLD voxel responses constitute the dependant variables and the design matrix contains the experimental manipulations. Fitting this model is slightly different to the univariate case, as covariance across the dependant variables (as in the case of the spatially near voxels in the searchlight) may lead to elliptical distributions of a class of data points in the dependant variable space. Therefore, fitting the dependant variables given a class membership might need to use different weights for different parameters relating to different dependant variables. In cvMANOVA, this is solved by using the Mahalanobis distance, a statistic that allows for each class distribution to have different covariance structure in the dependant variable space. This statistic is used to differentiate the hypothesis (i.e. the full model) multivariate distribution from the null multivariate distribution (i.e. the reduced model). In order to ensure that the unexplained variance does not contain any structure associated with the experimental variables, it is important to include all explanatory variables in the regression model. In this study, we apply this method to test models containing proposed integrated decision values, differential encoding of decision attributes and the stability of these attributes across conditions, a process often known as cross-decoding.

C.2 Methods

C.2.1 Experiment

Nineteen subjects took part in the fMRI study, of which sixteen were used for the analysis (one subject did not complete the study and two subjects had no variation in behaviour, meaning behavioural models could not be fit to their data). Prior to scanning, subjects were told that they had received an endowment of 100 US dollars before starting the experiment. They then read the experiment instructions and completed a pre-scan rating task in which they rated charities and household items by how much they were willing to pay (WTP) for them from their endowment. WTPs were elicited using a Becker de Groot auction which had been carefully explained in the instructions. In the scanner, subjects were presented with the same images that they had rated, in a random order. The image was presented alongside a label describing the charity or product and a price that they would need to pay in order to receive the item. Prices were selected from a normal distribution around the mean WTPs from a previous pilot data set using the same stimuli. Subjects then read the instructions for the fMRI experiment and began 4 runs of the fMRI task while undergoing fMRI scanning.

C.2.2 fMRI experiment

The functional imaging was conducted using a Siemens (Erlangen, Germany) 3.0 Tesla Trio MRI scanner to acquire gradient echo T2*-weighted echoplanar (EPI) images with BOLD contrast. Slices were oriented at 30 to the anterior commissureposterior commissure line. A Siemens eight-channel phased array coil was used to increase the BOLD signal. Each volume comprised 40 axial slices collected in an ascending manner. Data were collected in four sessions where the length of each session 586 volumes (24.4 min). The imaging parameters were as follows: echo time, 30ms; field of view, 192mm; in-plane resolution and slice thickness, 3mm; slice gap 0.3mm; repetition time, 2.5s. Whole-brain high-resolution T1-weighted structural scans (1x1x1 mm) were acquired from the 19 subjects and coregistered with their mean EPI images and averaged together to permit anatomical localization of the functional activations at the group level.

C.2.3 fMRI preprocessing

Image analysis was performed using SPM8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Images were motion corrected with realignment to the mean volume, spatially normalized to the standard Montreal Neurological Institute EPI template. Intensity normalization and high-pass temporal filtering (using a filter width of 128 s) were also applied to the data.

C.2.4 fMRI GLMs

In order to address the effects of an integrated value representation including the effect of the subject choice, a design matrix: GLM-1 was used. This included a parametric regressor was derived taking the (WTP - price) conditional on accepting the price (i.e. net value gained) and (price-WTP) conditional on refusing the item (i.e. net value lost). This regressor had onsets and durations aligned with the corresponding decision period as well as a boxcar regressor also with these onsets and durations.

In addition, in order to test the different ways that decision features were encoded for each charity and purchase choice, we created a design matrix: GLM-2. This included the main effect, WTP and price. As such, the design matrix was constructed with two dummy boxcar regressors for the charity and product conditions, with an onset for each purchase screen with duration equal to the reaction time and then convolved with a canonical haemodynamic response function. In addition, for each condition two more boxcar regressors were added with their amplitude modulated by the parameters for WTP and price. All GLMs included the motion regressors and dummy variables for the onset and duration of wach scan session. These GLMs were estimated on individuals' warped brains using SPM12.

These GLMs were used in a group level non-parametric analysis using the FSL (Winkler et al. 2014) randomise function combining threshold free cluster enhancement (Smith and Nichols 2009) with 5000 permutations of the data to determine a null distribution for voxel-wise multiple comparisons correction at the whole brain level.

GLM-1 and GLM-2 were also used for subsequent cvMANOVA analyses. The searchlight radius was set to 3 voxels (124 voxels, volume 3.68cm3). As

described in (Allefeld and Haynes 2014), for each subject, each of the four runs was tested using parameters derived from a training dataset consisting the remaining three runs. The average of these four folds was taken to be the D test statistic and data were prepared for permutation testing by sign flipping the training data (three runs leading to one true training data and seven sign flipped training data for building a null distribution) for each subject. These D-stat images were passed through a threshold free cluster enhancement process (Smith and Nichols 2009) implemented via the MatlabTFCE package and the resulting enhanced cluster images were used for all further analysis. Null hypothesis testing was achieved using the permuted maximum statistic method (Nichols and Holmes 2002). The null distribution was built over 5000 iterations by randomly selecting one D-stat image based on the sign flipped datasets per subject, taking the average across subjects and taking the maximum statistic in the resulting image per iteration. statistical significance was derived from comparing the average true D-stat against the null distribution.

C.3 Results

C.3.1 Behaviour

Prior to scanning, subjects completed a BDM auction to elict their WTP for the products and charities that would appear in the experiment while they were being scanned. The WTPs of charities and products were trend-level significant (product WTP: 14.2 ±6.6 mean/SD, charity WTP: 8.8 ±8.2 mean/SD, 2 sample t-test p=0.050). However, a similar trend was observed in the pilot data and prices were chosen as to offset the impact of this difference. As such, during scanning, the average difficulty (defined as —WTP-price—) of product and charity trials was not significantly different (product difficulty: 16.3 ±2.8 mean/SD, charity difficulty: 15.2 ±5.2 mean/SD, 2 sample t-test p=0.455). This was achieved despite the prices being not significantly different (product prices: 22.9 ±0.57 mean/SD, charity prices: 20.5 ±4.56 mean/SD, 2 sample t-test p=0.551). As expected, given that the average prices were higher than the average willingness to pay, subjects accepted the price only 33.5% of the time and this did not differ significantly betweeen the conditions (product buying fraction: 0.36 ±0.14 mean/SD, charity buying fraction: 0.31 ±0.29 mean/SD, 2 sample t-test p=0.618). Finally, subjects responded to price in a utility maximising manner based on their WTP 83%

of the time, significantly above 50% chance (one sample t-test p=0.00292 against a mean of 50%) and this was not significantly different between conditions (product percentage correct: 83.5 ±4.3 mean/SD, charity percentage correct: 82.6 ±13.6 mean/SD, 2 sample t-test p=0.810).

C.3.2 fMRI

Based on GLM-1, the integrated decision value was tested against the BOLD signal using mass univariate (with fsl's randomise) and multivariate (with the cvMANOVA framework) analyses. The univariate approach showed the BOLD signal in anterior cingulate cortex correlated negatively with the integrated decision variable, but only in product trials (table C.1). No significant effects were seen in the difference of decision value correlation between charity and product trials. In order to test whether there was a locally distributed representation, GLM-1 was also tested with cvMANOVA. First, the integrated decision value was tested across all trials, to examine whether cvMANOVA was sensitive to its effect on BOLD. Significant pattern discriminability was seen in large areas of the cortex, suggesting that locally distributed processing in these areas can track the effect of the integrated decision value (figure C.3, table C.2). In addition, the difference in this decision value between charity and product trials was tested with cvMANOVA, revealing a significant voxels in areas on the left lateral frontal cortex (figure C.4, table C.3), demonstrating that this value may be encoded differentially in these brain areas depending on the type of decision made.

Using GLM-2, the integrated decision value was broken up into price and WTP, allowing for these attributes and their interactions to be tested against the BOLD data. Similarly to the analysis approach with GLM-1, both univariate and multivariate analyses were tested. Similar to GLM-1, main effects and differences did not show significant effects except for a negtive correlation with price in the product condition alone (table C.4). However, these significant voxels were largely in precuneous and postcentral gyrus, suggesting that the individual contribution of price in this analysis may be a different effect to the integrated decision value seen in anterior cingulate. cvMANOVA was also applied to GLM-2, using a 2x2 MANOVA framework, where the voxels in the searchlight were predicted by two levels of condition and two levels of decision attribute. This decision attribute factor was tested against the BOLD signal across the brain and showed left lateral frontal cortex (figure C.5, table C.5), although in different areas to

the integrated decision result shown in figure C.4. One of the strengths of the MANOVA framework is that it allows cross decoding where there may be non-linear separation in the patterns exhibited. In GLM-2, we were able to test the stability of the effect of the attribute factor while accounting for the effects of the condition factor, i.e. decoding the attribute encoding across conditions. Figure C.6 (table C.6) shows significant voxels in the anterior cingulate cortex where the pattern of the attribute factor is stably discriminated.

C.4 Discussion

In this paper, we presented a fMRI study where subjects made decisions in charitable and product purchase contexts on whether to donate or buy at a given price. The act of purchasing captures one of the most important forms of decision making, in the formal exchange of personal resources for a good. This is also a linchpin of market economics for establishing how much a good is worth and what the correct price should be. This experiment examined the two key components of this process, in the willingness to pay and the price of a good. Although people can reveal their preferences through the decisions they make, understanding the mechanisms underlying these decisions can help form better inferences in novel situations. The charitable and product contexts of this experiment sought to explore the common and distinct mechanisms at play during purchase decisions. These mechanisms were assessed by testing BOLD correlates of these parameters during the decisions. A standard univariate approach showed that product decisions led to more activity for cheaper offers in the anterior cingulate and cuneous for the integrated decision value and the price respectively. The anterior cingulate has been implicated in several value-related functions (Kolling et al. 2016), and may represent some aspect of integrated value. While these results are not as strong as previous literature (Knutson et al. 2007), this may be due to power issues - we entered 16 subjects into the analysis compared to 26 in that study. The cuneous result is also often seen in the reward literature (Bartra et al. 2013), although more anterior to what was seen in this experiment. Again, this may be a power issue.

Considering the diverse range and complexity of the stimuli presented to subjects, we hypothesised that neural substrates may be expressed in spatially local but distinct areas. To this end a multivariate searchlight analysis (cvMANOVA) was employed to capture the local patterns relating to the

purchase decisions. First, this was used to assess the choice dependent value of subjects choices that was common across both contexts. This approach provided evidence that the integrated value is encoded in large areas including dorsomedial prefrontal cortex and left premotor cortex. Considering that subjects responded with their right hand for all trials, this is effect may relate to the value signal of a neural drift diffusion model as seen previously (Hare et al. 2011b). The distinctness of this choice dependant value between charitable and product contexts was also examined, with effects seen in lateral frontal cortex. In value based decision making, these regions have been associated with self-control (Hare et al. 2009) and strategy (Spitzer et al. 2007). Since the BOLD pattern is related to the choice value in this case, it may be that different patterns of self control are deployed when assessing charitable and product stimuli. This may indicate that other-orientated value and social norms exert different patterns of activity to a simple assessment of worth, but further work would be required to support this claim.

In order to more directly examine the attributes of the stimuli independent of choice, the cvMANOVA framework was applied to the WTP and price attributes without accounting for the choice outcome. This revealed a frontal lateral pattern, similar to that seen in the integrated value discriminating between contexts, but more anterior. If this pattern also relates to self control, then there may be multiple levels of inhibitory mechanisms at play during purchase decisions, some relating to the choice, but others relating to the stimuli themselves. In addition, a common pattern was found across charitable and product contexts in dorsal anterior cingulate that discriminated between the decision attributes. This area has been implicated in a wide range of value-related behaviors, in particular those involving a degree of conflict (Kolling et al. 2016), such as is seen in a purchase decision. This effect was seen to be stable in a cross decoding analysis where the pattern based on the data from one context maps can be used in the other context. However, this effect was not seen in the interaction between condition and decision attributes, so it is not likely to be related to different expressions of the decision conflict between condition. Instead, it may relate to the tradeoff between WTP and price.

It should be noted that previous fMRI experiments in charitable donations have often found regions associated with value representation such as ventral striatum (Moll et al. 2006) to be present. However, it should be noted that in the above study, the results were presented uncorrected for multiple comparisons and that the experiment was designed to test costly vs. costless donations. Since in our study the donation is always costly, this

may contribute to why that result is not present. Mechanisms of charitable donation have also been stuied more recently, where the mechanisms relating to empathy were examined. (Tusche et al. 2016). This found brain regions more often related to mentalizing than value. While we also show a role for anterior insula similar to theirs, our results relate to an integrated choice value and stimulus attribute measure for both charity and products, whereas Tusche et al. explained it as a measure of empathy for the charity solely in the charitable domain. This would need to be tested in purchase decisions in order to better distinguish the actual role of anterior insula in these value decisions.

In sum, this study presents a development in understanding choice attributes relating to decision making in charitable and personal contexts. While this provides evidence for the inclusion of medial and lateral frontal cortex, dorsal anterior cingulate and anterior insula, more research is required to more fully relate these effects to psychological phenomena. Further work exploring the empathetic and normative contributions to charitable giving might yield insights into how these attributes compare to similar strategic concerns and social norms that affect other domains of value based decision making.

C.5 References

Allefeld, Carsten and John-Dylan Haynes (Apr. 2014). "Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA". In: *NeuroImage* 89, pp. 345–357. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.11.043. URL: http://www.sciencedirect.com/science/article/pii/S1053811913011920.

Andreoni, James (1990). "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving". In: *The Economic Journal* 100.401, pp. 464–477. ISSN: 0013-0133. DOI: 10.2307/2234133. URL: http://www.jstor.org/stable/2234133.

Bartra, Oscar et al. (Aug. 2013). "The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value". In: *NeuroImage* 76, pp. 412–427. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.02.063. URL: http://www.sciencedirect.com/science/article/pii/S1053811913002188.

Batson, C. Daniel and Tecia Moran (Nov. 1999). "Empathy-induced altruism in a prisoner's dilemma". en. In: *European Journal of Social Psychology* 29.7, pp. 909–924. ISSN: 1099-0992. DOI: 10.1002/(SICI)1099-0992(199911)29:7<909::AID-EJSP965>3.0.CO;2-L. URL: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-0992(199911)29:7<909::AID-EJSP965>3.0.CO;2-L/abstract.

Beesly, Edward Spencer (1875). *System of Positive Polity*. en. Google-Books-ID: SQ3TAAAAMAAJ. Longmans, Green and Company.

Berkowitz, Leonard (Jan. 1972). "Social Norms, Feelings, and Other Factors Affecting Helping and Altruism11The author's research reported in this paper was carried out under grants from the National Science Foundation." In: *Advances in Experimental Social Psychology* 6, pp. 63–108. ISSN: 0065-2601. DOI: 10.1016/S0065-2601(08)60025-8. URL: http://www.sciencedirect.com/science/article/pii/S0065260108600258.

Clithero, John A. and Antonio Rangel (Sept. 2014). "Informatic parcellation of the network involved in the computation of subjective value". In: *Social Cognitive and Affective Neuroscience* 9.9, pp. 1289–1302. ISSN: 1749-5016. DOI: 10.1093/scan/nst106. URL: https://academic.oup.com/scan/article/9/9/1289/1675099/Informatic-parcellation-of-the-network-involved-in (visited on 08/05/2017).

Feigin, Svetlana et al. (Oct. 2014). "Theories of human altruism: a systematic review". In: *Annals of Neuroscience and Psychology*. URL: http://www.vipoa.org/neuropsychol/1/1/ (visited on 08/30/2017).

Gino, Francesca et al. (2016). "Motivated Bayesians: Feeling Moral While Acting Egoistically". In: *Journal of Economic Perspectives* 30.3, pp. 189–

212. URL: http://econpapers.repec.org/article/aeajecper/v_3a30_3ay_3a2016_3ai_3a3_3ap_3a189-212.htm.

Glazer, Amihai and Kai A. Konrad (1996). "A Signaling Explanation for Charity". In: *The American Economic Review* 86.4, pp. 1019–1028. ISSN: 0002-8282. DOI: 10.2307/2118317. URL: http://www.jstor.org/stable/2118317.

Hare, Todd A. et al. (May 2009). "Self-control in decision-making involves modulation of the vmPFC valuation system". eng. In: *Science (New York, N.Y.)* 324.5927, pp. 646–648. ISSN: 1095-9203. DOI: 10.1126/science.1168450.

Hare, Todd A. et al. (Jan. 2010). "Value Computations in Ventral Medial Prefrontal Cortex during Charitable Decision Making Incorporate Input from Regions Involved in Social Cognition". en. In: *Journal of Neuroscience* 30.2, pp. 583–590. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.4089-09.2010. URL: http://www.jneurosci.org/content/30/2/583 (visited on 06/23/2017).

Hare, Todd A. et al. (Nov. 2011b). "Transformation of stimulus value signals into motor commands during simple choice". en. In: *Proceedings of the National Academy of Sciences* 108.44, pp. 18120–18125. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1109322108. URL: http://www.pnas.org/content/108/44/18120 (visited on 09/11/2017).

Holmes, John G. et al. (Mar. 2002). "Committing Altruism under the Cloak of Self-Interest: The Exchange Fiction". In: *Journal of Experimental Social Psychology* 38.2, pp. 144–151. ISSN: 0022-1031. DOI: 10.1006/jesp.2001.1494. URL: http://www.sciencedirect.com/science/article/pii/S0022103101914945.

Hubbard, Jason et al. (Oct. 2016). "A general benevolence dimension that links neural, psychological, economic, and life-span data on altruistic tendencies". eng. In: *Journal of Experimental Psychology. General* 145.10, pp. 1351–1358. ISSN: 1939-2222. DOI: 10.1037/xge0000209.

Izuma, Keise et al. (Mar. 2009). "Processing of the Incentive for Social Approval in the Ventral Striatum during Charitable Donation". In: *Journal of Cognitive Neuroscience* 22.4, pp. 621–631. ISSN: 0898-929X. DOI: 10.1162/jocn.2009.21228. URL: http://dx.doi.org/10.1162/jocn.2009.21228.

Kagel, John H. and Alvin E. Roth (Sept. 2016). *The Handbook of Experimental Economics, Volume 2: The Handbook of Experimental Economics*. en. Google-Books-ID: y4LRDAAAQBAJ. Princeton University Press. ISBN: 978-1-4008-8317-2.

Kahnt, Thorsten et al. (May 2011). "Decoding different roles for vmPFC and dlPFC in multi-attribute decision making". In: *NeuroImage*. Multivariate

Decoding and Brain Reading 56.2, pp. 709–715. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2010.05.058. URL: http://www.sciencedirect.com/science/article/pii/S1053811910007913.

Knutson, Brian et al. (Jan. 2007). "Neural predictors of purchases". In: *Neuron* 53.1, pp. 147–156. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2006.11.010. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1876732/.

Kolling, Nils et al. (Sept. 2016). "Value, search, persistence and model updating in anterior cingulate cortex". eng. In: *Nature Neuroscience* 19.10, pp. 1280–1285. ISSN: 1546-1726. DOI: 10.1038/nn.4382.

Kriegeskorte, Nikolaus et al. (Mar. 2006). "Information-based functional brain mapping". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.10, pp. 3863–3868. ISSN: 0027-8424. DOI: 10.1073/pnas.0600244103.

Mayo, John W. and Catherine H. Tinsley (June 2009). "Warm glow and charitable giving: Why the wealthy do not give more to charity?" In: *Journal of Economic Psychology* 30.3, pp. 490–499. ISSN: 0167-4870. DOI: 10.1016/j.joep.2008.06.001. URL: http://www.sciencedirect.com/science/article/pii/S016748700800069X.

Moll, Jorge et al. (Oct. 2006). "Human frontomesolimbic networks guide decisions about charitable donation". en. In: *Proceedings of the National Academy of Sciences* 103.42, pp. 15623–15628. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0604475103. URL: http://www.pnas.org/content/103/42/15623 (visited on 07/16/2017).

Nichols, Thomas E. and Andrew P. Holmes (Jan. 2002). "Nonparametric permutation tests for functional neuroimaging: a primer with examples". eng. In: *Human Brain Mapping* 15.1, pp. 1–25. ISSN: 1065-9471.

Schwartz, Shalom H. (Jan. 1977). "Normative Influences on Altruism11This work was supported by NSF Grant SOC 72-05417. I am indebted to L. Berkowitz, R. Dienstbier, H. Schuman, R. Simmons, and R. Tessler for their thoughtful comments on an early draft of this chapter." In: *Advances in Experimental Social Psychology* 10, pp. 221–279. ISSN: 0065-2601. DOI: 10.1016/S0065-2601(08)60358-5. URL: http://www.sciencedirect.com/science/article/pii/S0065260108603585.

Simon, Herbert A. (1993). "Altruism and Economics". In: *The American Economic Review* 83.2, pp. 156–161. ISSN: 0002-8282. DOI: 10.2307/2117657. URL: http://www.jstor.org/stable/2117657.

Smith, Stephen M. and Thomas E. Nichols (Jan. 2009). "Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference". eng. In: *NeuroImage* 44.1, pp. 83–98. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2008.03.061.

Spitzer, Manfred et al. (Oct. 2007). "The Neural Signature of Social Norm Compliance". In: *Neuron* 56.1, pp. 185–196. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2007.09.011. URL: http://www.sciencedirect.com/science/article/pii/S089662730700709X.

Strahilevitz, Michal (Jan. 1999). "The Effects of Product Type and Donation Magnitude on Willingness to Pay More for a Charity-Linked Brand". In: *Journal of Consumer Psychology*. Ethical Trade-Offs in Consumer Decision Making 8.3, pp. 215–241. ISSN: 1057-7408. DOI: 10.1207/s15327663jcp0803_02. URL: http://www.sciencedirect.com/science/article/pii/S1057740899703517.

Tusche, Anita et al. (Apr. 2016). "Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving". en. In: *Journal of Neuroscience* 36.17, pp. 4719–4732. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.3392-15.2016. URL: http://www.jneurosci.org/content/36/17/4719 (visited on 06/23/2017).

Winkler, Anderson M. et al. (May 2014). "Permutation inference for the general linear model". In: *NeuroImage* 92, pp. 381–397. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2014.01.060. URL: http://www.sciencedirect.com/science/article/pii/S1053811914000913.

C.6 Tables

Table C.1: Regions with a negative BOLD correlation with decision value in product trials

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	t
20	Anterior Cingulate Cortex	L/R	-3	12	45	7.61
7	Anterior Cingulate Cortex	L	-6	30	36	11.88

Peak coordinates (x,y,z) are listed in MNI space. t values are test statistics derived from 5000 permutations of the data. These are significant at p <0.05 whole brain family wise error corrected for multiple comparisons

Table C.2: Regions with a cvMANOVA pattern for decision value across both charity and product decisions

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	D
4191	Precentral gyrus	L	-36	3	36	1.5×10^{-5}
	Temporal pole	L	-45	15	-12	1.4×10^{-5}
	Frontal pole	R	21	51	27	1.6×10^{-5}
	Precentral gyrus	R	57	9	9	1.9×10^{-5}
	Dorsofrontal gyrus	R	15	12	60	1.2×10^{-5}
238	Cuneous / precuneous	L/R	0	-72	18	2.2×10^{-5}
44	Postcentral gyrus	R	51	-15	57	1.8×10^{-5}
26	Superior temporal gyrus	R	57	-57	6	1.9×10^{-5}
5	Occipital cortex	R	54	-63	-12	2.3×10^{-5}
4	Cerebral white matter	R	27	27	24	1.4×10^{-5}
4	Frontal pole	R	30	42	48	1.6×10^{-5}
4	Frontal pole	R	21	48	45	1.2×10^{-5}

Peak coordinates (x,y,z) are listed in MNI space. D values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at p <0.05 whole brain family wise error corrected for multiple comparisons

Table C.3: Regions with a cvMANOVA pattern that differentiates decision value encoding in charity from product decisions

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	D
106	Frontal pole	L	-42	42	24	$2.4 \times 10^{-}$
66	Precentral gyrus	L	-42	3	45	$1.8 \times 10^{-}$
58	Dorsomedial prefrontal cortex	L	-24	39	48	$1.9 \times 10^{-}$
37	Precentral gyrus	L	-54	-12	45	$2.4 \times 10^{-}$
31	Supracalcine cortex	L/R	3	-81	9	$1.9 \times 10^{-}$
13	Medial frontal cortex	L	-33	24	27	$1.7 \times 10^{-}$
4	Occipital cortex	L	-27	-96	12	$2.6 \times 10^{-}$
1	Precentral gyrus	L	-48	0	54	$1.5 \times 10^{-}$

Peak coordinates (x,y,z) are listed in MNI space. D values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at p <0.05 whole brain family wise error corrected for multiple comparisons

Table C.4: Regions correlating BOLD with price in product trials more than charity trials

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	t
18	Postcentral gyrus	L	-42	-33	54	5.31
4	Precuneous cortex	L/R	0	-60	45	4.74
2	Precuneous cortex	L/R	-3	-69	51	4.62
1	Postcentral gyrus	R	42	-33	48	6.31
1	Angular gyrus	L	-30	-60	57	6.07
1	Occipital cortex	L	-21	-66	60	5.98

Peak coordinates (x,y,z) are listed in MNI space. D values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at p <0.05 whole brain family wise error corrected for multiple comparisons

Table C.5: Regions with a cvMANOVA pattern differentiating price and WTP for both charity and product decisions

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	D
299	Medial frontal gyrus	L	-42	3	36	$4.1 \times 10^{-}$
35	Frontal pole	L	-24	42	48	$4.2 \times 10^{-}$
24	Fusiform gyrus	L	-12	-75	-18	$4.2 \times 10^{-}$
7	Dorsomedial prefrontal cortex	L/R	3	48	48	$4.2 \times 10^{-}$
1	Precuneous cortex	L/R	-6	-72	36	$4.8 \times 10^{-}$

Peak coordinates (x,y,z) are listed in MNI space. D values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at p <0.05 whole brain family wise error corrected for multiple comparisons

Table C.6: Regions with a cvMANOVA pattern of price and WTP that is stable across charity and product decisions

Extent	Region	Hemisphere	x(mm)	y(mm)	z(mm)	D
6	Dorsal anterior cingulate	L/R	0	24	36	5.8×10^{-5}

Peak coordinates (x,y,z) are listed in MNI space. D values are test statistics derived from 5000 permutations of the data. All regions are local cluster peaks at least 50mm apart. These are significant at p <0.05 whole brain family wise error corrected for multiple comparisons

C.7 Figure Legends

Figure C.1: Subjects completed a buy / no buy task for charitable donations and products in randomized order. The durations of each screen are shown below each screenshot.

Figure C.2: Behavioural analysis showed no significant difference in behaviour between the charity and product decisions, except for a in willingness to pay (WTP), although the difficulty of the decisions (as the unsigned difference between price and WTP) did not differ at time of decision.

Figure C.3: Pattern discriminablity of the integrated value of a decision across all trials showed large areas of cortex, particularly in fronto-medial, insula and cuneous regions. Image shown at MNI coordinates $(1, -26, 14)$.

Figure C.4 Pattern discriminablity of the difference in integrated value between charity and oroduct decisions. Areas of left lateral prefrontal, left lateral frontal cortex amd left precentral cortex showed significant pattern discriminablity. Image shown at MNI coordinates $(-44, 39, 18)$.

Figure C.5 Pattern discriminablity of the difference between willingness to pay and price correlations across all trials. areas of left lateral prefrontal and left lateral frontal cortex showed significant pattern discriminablity. Image shown at MNI coordinates $(-38, 10, 44)$.

Figure C.6 Pattern stability of the difference in willingness to pay and price correlations, after taking out the main effects of trial condition. The anterior cingulate pattern is shown to be stable (i.e. anterior cingulate can cross-decode WTP and price across charity and product decisions). Image shown at MNI coordinates $(1, 14, 18)$.

C.8 Figures

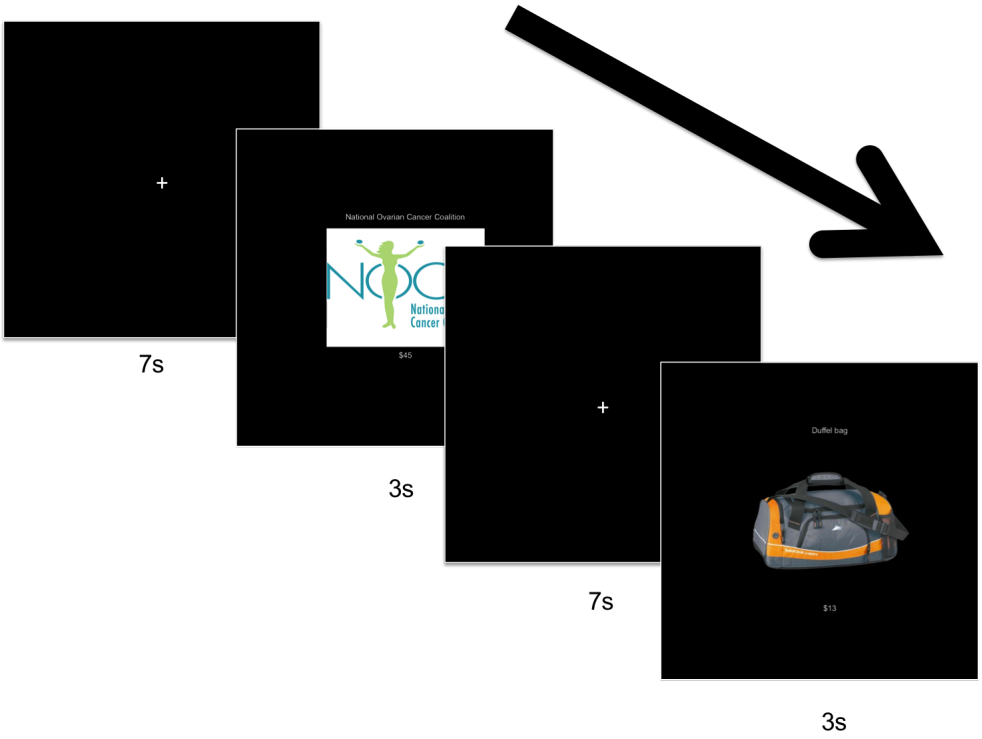

Figure C.1: Task Design

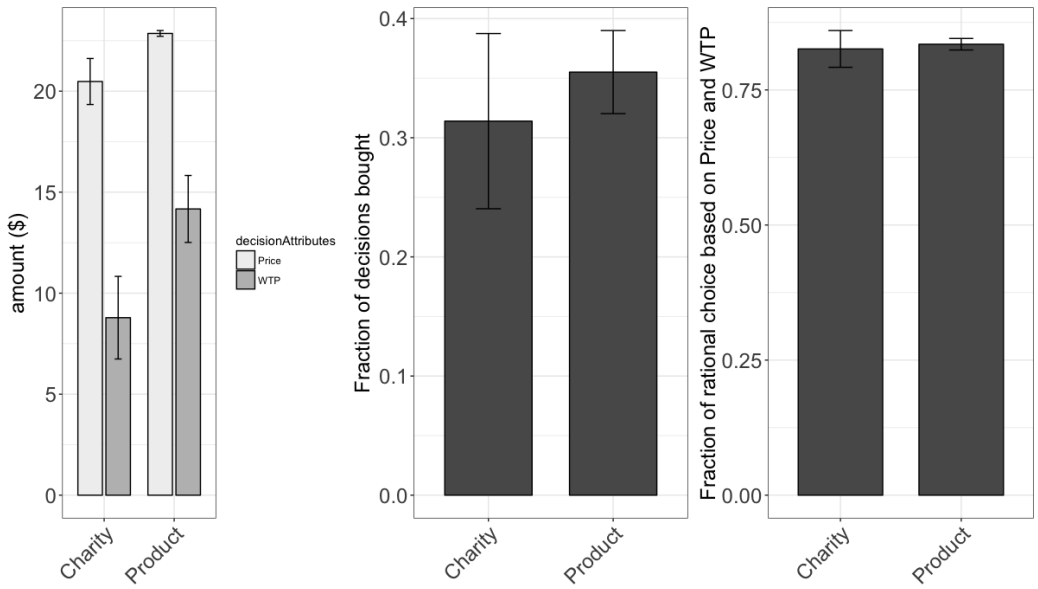

Figure C.2: Behavioural data by condition

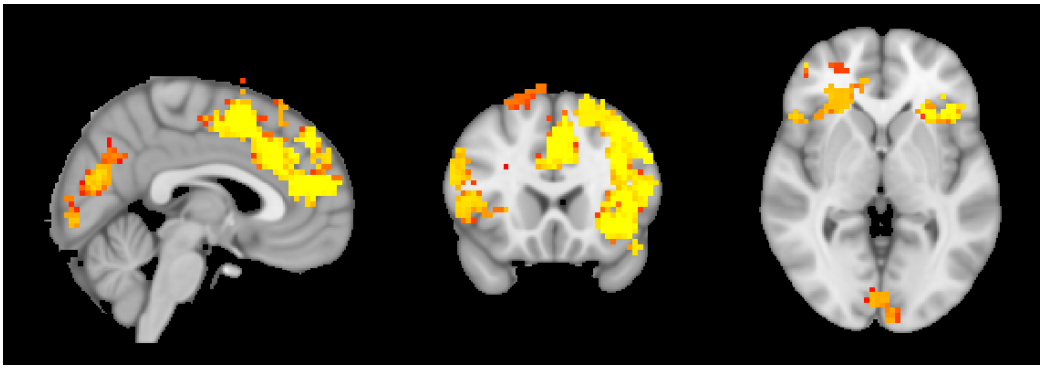

Figure C.3: cvMANOVA decision value in charity and product conditions

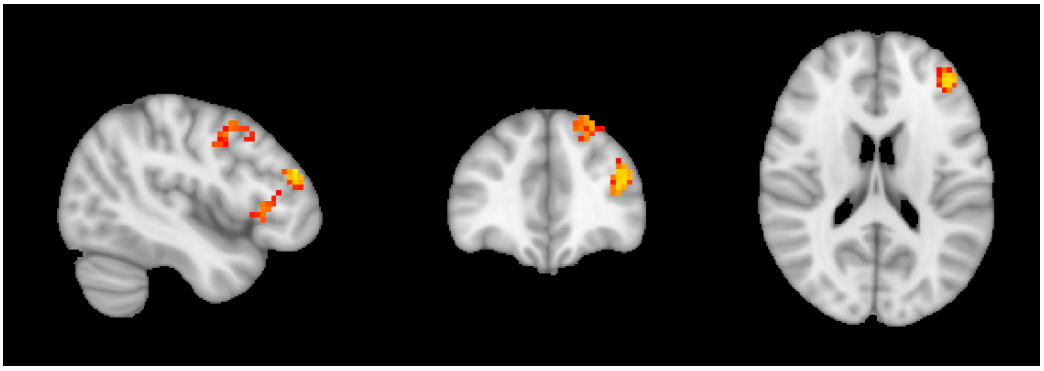

Figure C.4: cvMANOVA differential decision value between charity and product conditions

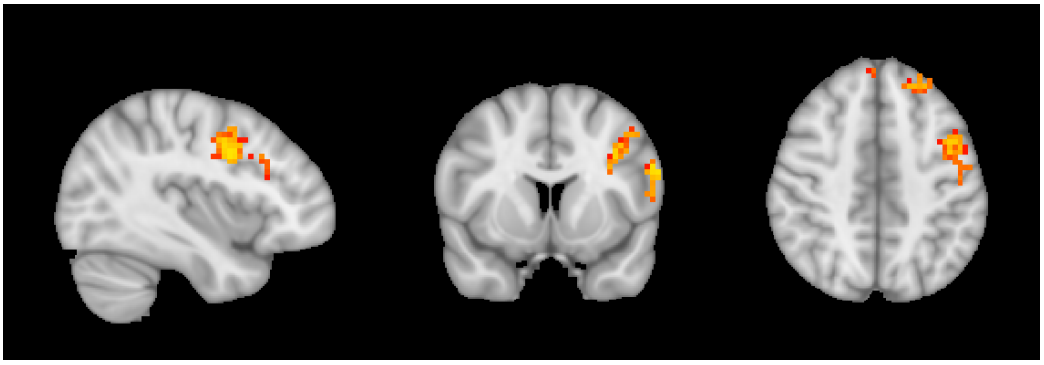

Figure C.5: cvMANOVA main effect of WTP vs price for both conditions

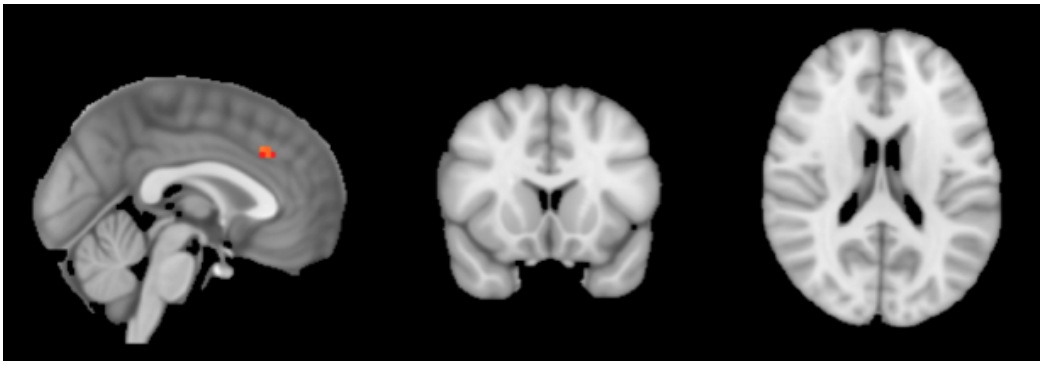

Figure C.6: cvMANOVA stability of WTP vs price across both conditions

Curriculum Vitae

Personal details

Aidan Makwana

Date of Birth:	07 FEBRUARY 1986

Education

September 11 – February 18	Doctoral program in Neuroeconomics University of Zurich, Department of Economics
September 04 – July 08	Master of Engineering in chemical engineering, Manchester university, Department of chemical engineering

Professional experience

February 09 – August 11	Imaging assistant scientist, GlaxoSmithKline, London